# Studying Depression through Big Data Analytics on Twitter

Angela Leis Machín

DOCTORAL THESIS UPF / 2020

THESIS DIRECTOR

Dr. Ferran Sanz

DEPARTMENT OF EXPERIMENTAL AND HEALTH SCIENCES

*Para Miguel Ángel y Elena*

## Acknowledgements

En el momento que comencé esta aventura del doctorado, me encontraba en una de esas encrucijadas en las que la vida a veces nos pone a prueba. Por este motivo, esta tesis además de ser el resultado final del trabajo realizado a lo largo de estos años es muy especial para mí, ya que ha supuesto un gran reto tanto profesional como personal, pero también me ha permitido conocerme mejor y saber que no hay límites si no quieres, nunca es tarde para seguir aprendiendo.

En primer lugar, me gustaría agradecer a mi director de tesis, el profesor Ferran Sanz, la oportunidad de haber podido desarrollar esta tesis en la que hemos explorado el mundo de las Redes Sociales desde un punto de vista biomédico. Desde el primer momento, tanto su confianza en el éxito de este proyecto, su apoyo continuo, la capacidad de encontrar siempre el momento para resolver mis dudas, su cercanía, su energía y entusiasmo han sido un verdadero ejemplo para mí tanto profesional como personalmente.

También agradecer a Laura Furlong su apoyo durante estos años y siempre dispuesta a darme sus consejos y ofrecerme palabras de apoyo. Asimismo, me gustaría dar las gracias a todos los compañeros con los que he tenido el placer de coincidir a lo largo de estos años, a Alexia, Alba, Àlex, Emilio, Emre, Janet, Judith, Josep, Juanma, Lluís, Francesco, Miguel Ángel y Núria, y que me han ayudado con sus aportaciones a seguir mejorando. Esta tesis es un ejemplo más de la importancia de la colaboración multidisciplinar en

Informática Biomédica y que caracteriza especialmente nuestro grupo. En especial, quiero mencionar la inestimable ayuda y soporte de Francesco. Muchas gracias Francesco por tu paciencia estos años y recuerdo con cariño la lucha con los múltiples análisis que realizamos. Esta tesis tampoco hubiera sido posible sin las aportaciones desde su experiencia médica de Miguel Ángel.

Muchas gracias a Carina, Chus y Martina, por poder contar siempre con vuestro apoyo. Y a Alfons y Miguel Ángel Sánchez por su soporte técnico.

A lo largo de estos años muchas personas han contribuido a que esta tesis sea una realidad. Muchas gracias a Olga Valverde, Marta Torrens y Claudio Castillo por su ayuda e implicación desde el primer momento. A todos los profesionales del CAS Barceloneta, CAS Fòrum y CAS Santa Coloma por su cálida acogida, pero en especial a Joana y Francina, y qué hubiera sido de mí sin Carme. Una especial mención para todos los pacientes que compartieron conmigo sus vivencias en la consulta. Igualmente quería agradecer su colaboración a los psiquiatras y psicólogos del Instituto de Neuropsiquiatría y Adicciones (INAD), así como a Jesús Pujol, Paco Díaz y Ramón Sarrias. Muchas gracias también, a todos los estudiantes de la Universitat Pompeu Fabra que participaron.

A Martin, muchas gracias por tu paciencia y buenos consejos para explorar las palabras más adecuadas en cada ocasión.

Esta aventura no habría sido posible si hace años no hubiera seguido mi corazón, dejando atrás mi querida isla. A mi marido Miguel Ángel, haces que cada día sea especial y único, ¡una vida a tu lado no es suficiente! Muchas gracias por tu cariño, apoyo y paciencia. Por estar a mi lado en esos momentos en los que no sabía como continuar y por creer siempre en mí, animándome a seguir adelante. A mi hija Elena, que durante todo el verano antes del depósito de la tesis, todos los días preguntaba porque tardaba tanto en escribirla y el día antes de imprimirla, preguntó… pero ¿para qué sirve la tesis? No sé si habrá entendido del todo para qué sirve, pero estoy segura de que sabrá ver que el esfuerzo y la constancia tienen su recompensa. A Àlex y Pau, que durante estos años me han visto siempre trabajando y luchando para llegar hasta aquí.

A mis queridos padres, Manolo y Angelines, que siempre me han apoyado en mis decisiones y han confiado en mí, y a mi hermano Manuel. Me hubiera encantado poder compartir este momento con mi padre que siempre decía que mi fuerza de voluntad me llevaría lejos.

Por último, quería dedicar esta tesis a todas aquellas personas que sufren depresión, y que en estos momentos tan complicados que estamos viviendo con la COVID-19, se les hace todavía más difícil el día a día. También a todos los profesionales de la salud que siguen trabajando incansablemente en las actuales circunstancias desde todos los ámbitos, tanto clínico como de investigación.

## Abstract

Mental disorders have become a major concern in public health, since they are one of the main causes of the overall disease burden worldwide. Depressive disorders are the most common mental illnesses, and they constitute the leading cause of disability worldwide. Language is one of the main tools on which mental health professionals base their understanding of human beings and their feelings, as it provides essential information for diagnosing and monitoring patients suffering from mental disorders.

In parallel, social media platforms such as Twitter, allow us to observe the activity, thoughts and feelings of people's daily lives, including those of patients suffering from mental disorders such as depression. Based on the characteristics and linguistic features of the tweets, it is possible to identify signs of depression among Twitter users. Moreover, the effect of antidepressant treatments can be linked to changes in the features of the tweets posted by depressive users.

The analysis of this huge volume and diversity of data, the so-called "Big Data", can provide relevant information about the course of mental disorders and the treatments these patients are receiving, which allows us to detect, monitor and predict depressive disorders.

This thesis presents different studies carried out on Twitter data in the Spanish language, with the aim of detecting behavioral and linguistic patterns associated to depression, which can constitute the basis of new and complementary tools for the diagnose and follow-up of patients suffering from this disease.

## Resumen

Los trastornos mentales constituyen un problema de la salud pública, siendo una de las principales causas de morbilidad en todo el mundo. Los trastornos depresivos son las enfermedades mentales más comunes y constituyen la mayor causa de discapacidad en el mundo. El lenguaje es una de las principales herramientas en el que los profesionales de la salud mental se basan para comprender al ser humano y sus sentimientos, proporcionando información esencial para diagnosticar y monitorizar a los pacientes que padecen trastornos mentales.

Las redes sociales, como Twitter, nos permiten conocer la actividad, pensamientos y sentimientos de las personas, incluyendo las que padecen depresión. A partir de las características y rasgos lingüísticos de sus tweets, es posible identificar signos de depresión entre los usuarios de Twitter. Además, el efecto de los tratamientos antidepresivos se puede relacionar con cambios en las características de los tweets de usuarios depresivos. El análisis de este gran volumen y diversidad de datos, "Big Data", puede aportar información relevante sobre la evolución de los trastornos mentales y los tratamientos que reciben estos pacientes, permitiendo detectar, monitorizar y predecir los trastornos depresivos.

Esta tesis presenta diferentes estudios realizados sobre datos de Twitter en español, con el objetivo de detectar patrones conductuales y lingüísticos asociados a la depresión, que pueden constituir la base de nuevas herramientas complementarias para el diagnóstico y seguimiento de los pacientes que padecen esta enfermedad.

## Preface

This thesis is focused on the study of depression by means of the analysis of social media posts written Spanish. There are no studies in this language, since up to mow they have been usually focused on messages written in English. In addition, the language used in the posts use not to be medical or expert based. This means that it was necessary to create new approaches for the selection of words related to depression and for making the most of several analytical tools to be used.

Due to the complexity of depression and the extraordinary diversity in the way people express themselves when describing their illness, it was critical to perform a thorough analysis of the specific language used by patients suffering from this condition, before deciding the best words to be used for detecting tweets related to this disease. In addition, grammatical gender forms should be considered when analyzing some languages such as Spanish. For this reason, this thesis incorporates a comprehensive revision and collection of Spanish words commonly used by patients suffering from depression, which is freely available as an open resource for research purposes (GitHub), as well as the use of diverse linguistic tools for a thorough analysis of the features of tweets.

# Table of Contents

## List of figures

## List of tables

# 1. INTRODUCTION

*"Nothing in life is to be feared, it is only to be understood. Now is the time to understand more, so that we may fear less"*

Marie Curie

## 1.1. Mental health

### 1.1.1. Mental health concept

Mental health is an essential component of our health at every stage of life, from childhood and adolescence through adulthood. The World Health Organization (WHO) defines mental health as a "state of well-being in which people realize their potential, cope with the normal stresses of life, work productively, and contribute to their communities" (WHO, 2013). Good mental health is about being cognitive, emotionally and socially healthy, determining the way we think and feel in relation with others and how we make choices and manage stress. There are several factors, such as genetic, sociocultural, economic, political and environmental aspects, which shape and influence our mental health.

The WHO's definition entails a significant advance concerning the definition of mental health as a state of absence of mental illness. Nevertheless, considering that well-being is a key aspect of mental health, it is difficult to adapt it to the many challenging life situations, sometimes sad, disgusting, frightening or unsatisfactory (Galderisi et al., 2015; Huber et al., 2011).

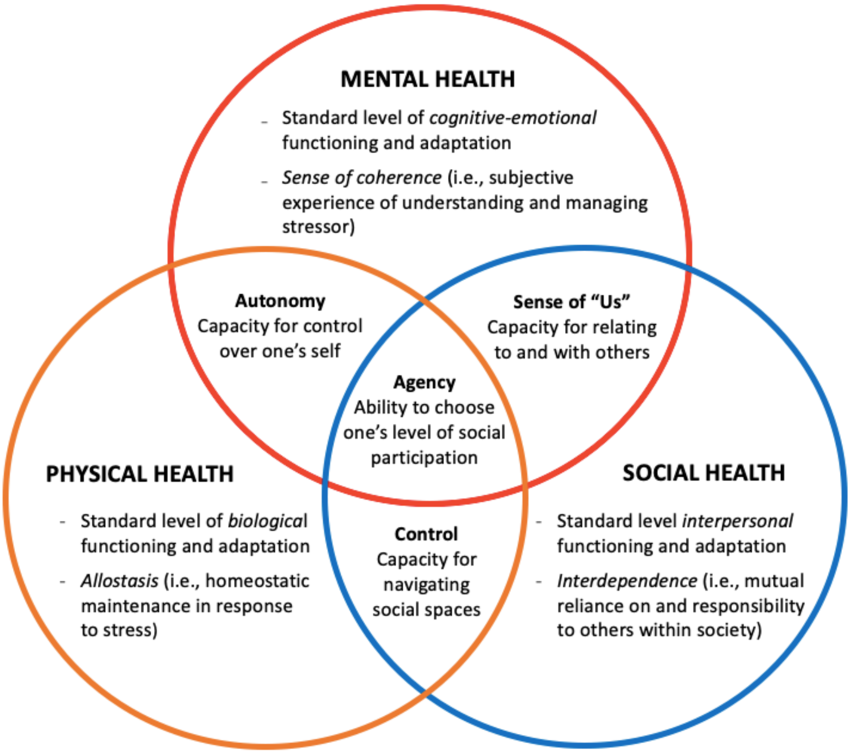
There is a lack of consensus on the definition of mental health that may have several implications for the development of research, for the application of health policies and in clinical practice (Manwell et al., 2015). For this reason, there is a need to reframe the concept in culturally sensitive and inclusive terms, and the criteria for defining mental health have to be empirically and longitudinally validated (Vaillant, 2012).

Galderisi et al. (2017) reconceptualize the definition of mental health as "a dynamic state of internal equilibrium which enables individuals to use their abilities in harmony with universal values of society". That involves the recognition and management of our own emotions, the capacity to cope with negative life events and adapting to the different social roles that we play. This flexibility allows to empathize with others and to find the internal equilibrium between the body and mind, both essential components of mental health. In this way, in the different periods of life people with good mental health may feel positive and negative human emotions but, at the same time, they have the ability to adapt to adverse conditions and to timely restore the dynamic state of their psychological balance. This ability is known as resilience (Galderisi et al., 2015).

Mental health is not an isolated dimension of our health and it is influenced by physical and social domains (Huber et al., 2011; WHO, 2013). Manwell et al. (2015) proposed a trans-domain model of health to develop a comprehensive definition for all aspects of health, including cross-cultural points of view and based on the aforementioned three domains of health. By expanding these domains and considering the dynamic areas of overlapping and synergy among them (i.e., autonomy, agency, control and sense of us), the concept of mental health can be more accurately defined as it is depicted in Figure 1.

**Figure 1**

*Transdomain model of health*




*Note.* Adapted from Manwell et al. (2015). What is mental health? Evidence towards a new definition from a mixed methods multidisciplinary international survey. *British Medical Journal Open*, *5*(6), e007079.

In the last few years, diverse mental health continuum models have emerged trying to explain the differences between a good and a poor mental health. These models are based on the premise that people can move along the continuum during their life. In 2008, Suldo and Shaffer (2008) created the first mental health model, characterized by a dichotomous approach which equates mental health to the absence of mental illness. This model describes 'health' and 'illness' as two separate states. Nevertheless, mental health is not just the absence of mental illness (Keyes, 2005), and Suldo's model does not explain why sometimes people may be in a status between health and illness, showing symptoms of stress and emotional distress but insufficient to diagnose a mental health condition (Bickman, 1996).

More recently, Keyes suggested a dual continuum model (2002). This model argues that mental health and mental illness are independent of each other and are not at the opposite side of the same continuum (Keyes, 2005). According to this model, mental health is a combination of emotional, psychological, and social well-being and distinguishes between the concept of "flourishing" to refer to people with good mental health and no mental illness, and "languishing" as those with severe stress on their mental health (Keyes, 2014). The main implication of the dual continuum model is that the absence of mental illness does not mean the presence of mental health. Nevertheless, the presence of a mental illness does not imply the absence of some level of good mental health (Keyes, 2014). The most controversial aspect of this model is that it raises that people may have symptoms associated with a mental health condition, but they can still be flourishing. However, some authors point out that this may only be applied to specific conditions such as chronic schizophrenia but not others, such as major depression (Huppert & So, 2013).

The different mental health continuum models are seeking to cope with the complexity and variability of the mental health concept from different approaches. In any case, it is essential to bear in mind that the mental health concept is characterized by being multidimensional, dynamic and changeable, as it reflects our thoughts, feelings, perceptions, behaviors, relationships and day-to-day functioning.

### 1.1.2. Mental disorders classification

As discussed in the previous section, mental health and illness are not binary states but a continuum, ranging from normal functioning to different types of functional impairment and clinical illnesses or disorders. The definition and classification of mental disorders are key questions for clinical, public health, and research purposes (Clark et al., 2017; Üstün & Ho, 2017). Mental illness, also called mental disorder, is a "recognized, medically diagnosable illness that results in the significant impairment of an individual's cognitive, affective, or relational abilities" (Epp, 1988). Different definitions of this concept have been proposed over the years and have been subject to debate.

The main classifications for the mental disorder are the International Classification of Diseases (ICD) (WHO, 2019a) and the Diagnostic and Statistical Manual of Mental Disorder (DSM) (American Psychological Association [APA], 2013). These classification systems and nomenclatures have evolved over time as scientific knowledge progresses and they provided us an overview of what we must understand as a mental disorder. In psychiatry, as other areas of medicine, the limits between normality and abnormality, based on science, change over time (Kendler, 2009; Stein, 2008; Stein et al., 2010). The DSM is used for research due to its operational criteria (Clark et al., 2017) while according to a survey conducted among nearly 5,000 psychiatrists from 44 countries, 70% of them usually applied the ICD in clinical environments, particularly in Europe (Reed et al., 2011).

In the International Classification of Diseases (ICD) 11th Revision, WHO defines a mental disorder as “a mental, behavioral or neurodevelopmental disorders are syndromes characterized by clinically significant disturbance in an individual's cognition, emotional regulation, or behavior that reflects a dysfunction in the psychological, biological, or developmental processes that underlie mental and behavioral functioning” (WHO, 2019b).

The American Psychiatric Association (APA) defined mental disorders in the Diagnostic and Statistical Manual of Mental Disorders (DSM-5), as “a syndrome characterized by clinically significant disturbance in an individual’s cognition, emotion regulation, or behavior that reflects a dysfunction in the psychological, biological, or developmental processes underlying mental functioning. Mental disorders are usually associated with significant distress or disability in social, occupational, or other important activities” (APA, 2013). The classification of DSM-5’s mental disorders is presented in Table 1.

**Table 1**

*High level classification of mental disorders according to DSM-5*

| ***DSM-5 Classification*** |
| --- |
| • Neurodevelopmental Disorders<br>• Schizophrenia Spectrum and Other Psychotic Disorders<br>• Bipolar and Related Disorders<br>• Depressive Disorders<br>• Anxiety Disorders<br>• Obsessive-Compulsive and Related Disorders<br>• Trauma- and Stressor-Related Disorders<br>• Dissociative Disorders<br>• Somatic Symptom and Related Disorders<br>• Feeding and eating Disorders<br>• Elimination Disorders<br>• Sleep-Wake Disorders<br>• Sexual Dysfunctions<br>• Gender Dysphoria<br>• Disruptive, Impulse-Control, and Conduct Disorders<br>• Substance-Related and Addictive Disorders |

| ***DSM-5 Classification*** |
| --- |
| • Neurocognitive Disorders<br>• Personality Disorders<br>• Paraphilic Disorders<br>• Other mental Disorders<br>• Medication-Induced Movement Disorders and Other Adverse Effects of Medication<br>• Other Conditions that may be a focus of Clinical Attention |

*Note.* Adapted from American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Association; 2013.

The causes of mental disorders are diverse and not fully understood. Many of these disorders are caused by a combination of biological, psychological and environmental factors that can contribute to the development of mental disorders (Arango et al., 2018). In the same way, the role of the social aspects is critical in the etiology and symptomatology for the most of these disorders (Patel et al., 2009; Patel et al., 2016). The main social aspects associated with several mental disorders are shown in Table 2.

**Table 2**

*Main social aspects associated with mental disorders*

| **Social Aspects of Mental Disorders** |
| --- |
| • Demographic factors (i.e., age, gender and ethnicity)<br>• Socioeconomic status such as low income, employment, income inequality, low education and low social support |

| Social Aspects of Mental Disorders |
| --- |
| • Neighborhood factors, including inadequate housing, overcrowding, neighborhood violence<br>• Environmental events such as natural disasters, conflict, climate change and migration<br>• Social changes in income, urbanization or environmental degradation |

*Note*. Adapted from Patel et al. (2009). Social determinants of mental disorders. Priority public health conditions: from learning to action on social determinants of health. Geneva: World Health Organization.

Mental disorders are very complex and heterogeneous conditions, which may vary depending on every individual in particular and the combination of several different factors (Clark et al., 2017). In general, they are characterized by combination of abnormal thoughts, perceptions, emotions, behavior and relationships with others (WHO, 2018a). These disorders are usually associated to significant distress or disability in personal, family, social, educational, occupational or other important activities (APA, 2013; WHO, 2019a). They have devastating consequences for both patients and their families and also affect the society through the magnitude of the economic impact of mental disorders (Global Burden of Disease Study 2013 Collaborators, 2015; Hu, 2004; Marcus et al., 2012; Trautmann et al., 2016; Patel et al., 2016; Whiteford, et al., 2013; Wongkoblap et al., 2017). This economic impact affects directly in the healthcare system, and indirectly on productivity and economic growth, even more than other chronic diseases such as cancer or diabetes. The global economic cost of mental disorders was around US$2.5 trillion in 2010, and it is estimated to double by 2030 (Trautmann, 2016). Moreover, the global impact in terms of economic loss is estimated to $16 trillion by 2030 (Bloom et al; 2011; Whiteford et al., 2013; Whiteford et al., 2016).

Likewise, mental conditions are associated with higher frequency of comorbidities and people with mental disorders endure a high burden of mortality at the individual and population levels (Murray & Lopez, 2013; Walker et al., 2015; Whiteford et al., 2013). Furthermore, there is a two-way relationship between mental disorders and unhealthy behaviors such as poor diet and physical lack of activity. These factors contribute to the co-occurrence of mental disorders with several physical comorbidities such as cancer, cardiovascular disease, diabetes, obesity, hypertension, musculoskeletal or respiratory tract diseases (Bloom et al., 2011; Meng et al., 2012; Mnookin, 2016). In addition, mental disorders magnify the probability of drug and alcohol abuse, which can involve risky sexual behaviors, such as transmitted infectious diseases and HIV and other injuries (Bloom et al., 2011). On the other hand, mental disorders have been associated to suicide. In fact, patients suffering from a mental disorder have an eight-fold increased risk of suicide compared to those who do not suffer from this disease (Bloom et al., 2011; San Too et al., 2019). Moreover, serious mental disorders such schizophrenia, bipolar disorder and major depressive disorder (MDD) have been related to a 60% higher chance of dying prematurely from non-communicable diseases that are disregarded because of the mental condition (Vigo et al., 2016). As a whole, mental disorders are the cause of 14.3% of deaths worldwide (8 million deaths per year) (Vigo et al., 2016). All of these circumstances have important effects on life expectancy (Charlson et al., 2015; Walker et al., 2015; Whiteford et al., 2013).

In this scenario, it is necessary to develop strategies for prevention and management aimed to detect and follow up mental disorders and their comorbidities, thus reducing unnatural deaths in this vulnerable population (Vigo et al., 2016).

### 1.1.3. Mental disorders prevalence

In the early 1990s, the first Global Burden of Disease study, known as GBD, confirmed that mental health and substance use disorders started to be a main concern in global health (Whiteford et al., 2013; Whiteford et al., 2016). This first GBD study pointed out that mental, neurological and substance use disorders represented more than a quarter of all non-fatal burden in terms of years lived with disability (YLDs) (Whiteford et al., 2013). In 2010, mental and substance use disorders represented the 10.4% of global daily disability adjusted life years (DALYs) burden of disease and turned into the leading cause of YLDs, liable for 22.9% of global YLDs, therefore the GBD adopted the decision of separating these disorders from the neurological ones for their estimations (Whiteford et al., 2015). Mental and substance use disorders were responsible for 7.4% of the global burden of diseases. The most disabling disorder was depression (Whiteford et al., 2015).

In the last few years, mental disorders have become a major concern in public health, and they are one of the main causes of the overall disease burden worldwide. (Global Burden of Disease Study 2013 Collaborators, 2013; Marcus et al., 2012; Patel et al., 2016; Trautmann et al., 2016; Whiteford et al., 2013; Wongkoblap et al., 2017). In the next years, the burden of mental health disorders will increase, due to the demographic changes such as the growth and aging of the world population, as well as the lack of access to healthcare and the limited resources (Baingana et al., 2015; Global Burden of Disease Study 2013 Collaborators, 2015; Sartorius et al., 2017). It is estimated that, in both low and middle-income countries, 80% of the people suffer an episode of a mental disorder in their lifetime. These disorders involve a great burden for society, representing almost one over three years lived with disability worldwide (Mnookin, 2016). In 2016, mental and addictive disorders accounted for 7% of all global burden of disease in DALYs and 19% of YLD, affecting more than 1 billion people globally. The burdens of mental disorders are frequently underestimated (Vigo et al., 2016) and around 10% of the world's population suffer from one or more mental disorders (Helliwell et al., 2012; Patel & Saxena, 2014). Among all the internalizing disorders such as depressive, bipolar anxiety and eating disorders (Krueger, 1999), depression was associated with most DALYs for both sexes, with higher rates in women (GBD 2015 DALYs and HALE Collaborators, 2016; Rehm, & Shield, 2019).

Nevertheless, despite these overwhelming situations regarding the amount of people suffering from mental and addictive disorders, as well as the personal and economic costs as, treatment rates are low (Rehm, & Shield, 2019). Meanwhile treatment is initiated many years after the disorder begins in developed countries, treatment gaps of mental health in developing countries was estimated to be around 90% (Whiteford et al., 2013), remaining unavailable or inadequate in most of the world (Kohn et al. 2004). In this way, there is a global treatment gap between those who need and those who received treatment for their mental disorders (between 32% and 78%, depending on the disorder). In low- and middle-income countries, between 76% and 85% of people with mental disorders do not receive treatment. In high-income countries, this percentage drops to 35% and 50% (Wang et al., 2007). Likewise, another reason that influences this mental health treatment gap is the lack of qualified mental health service providers in a major part of the world (Clark et al., 2017). As indicated by WHO, one in every ten people will need mental health care at some point in their life, however in low-income countries the rate of mental health workers is lower as 2 per 100.000 population, compared with more than 70 in high-income countries (WHO, 2018b).

Persons with mental illness are mostly neglected. The combination of widespread stigma and discrimination contribute to social exclusion and also to the disparity between the global burden of mental disorders, and the attention to these conditions (Kleinman, 2009; Vigo et al.; 2016; Vigo et al., 2019). Healthcare systems worldwide have not covered the enormous demand of mental disorders, nor do they have the necessary human and financial resources (WHO, 2018b). The burden of mental disorders is at least two-fold in the less developed countries as compared to the well-developed ones (Rozatkar, 2016). The healthcare systems spend significantly less resources in mental health than in other types of diseases, which is reflected in lower availability, accessibility and quality of services (Kleinman, 2009; Vigo et al., 2016; Vigo et al.; 2019), as well as in an inadequate prevention programming (Saxena et al., 2007).

In summary, the impact of mental illness on global health tends to be underestimated and for this reason it is crucial to support research to develop and implement better prevention and treatment options, as well as tailor cost-effective interventions culturally acceptable, with particular focus on low- and middle-income countries (Patell et al., 2018; Pathare et al., 2018; Rehm & Shield, 2019; Whiteford et al., 2015).

# 1.2. Depression

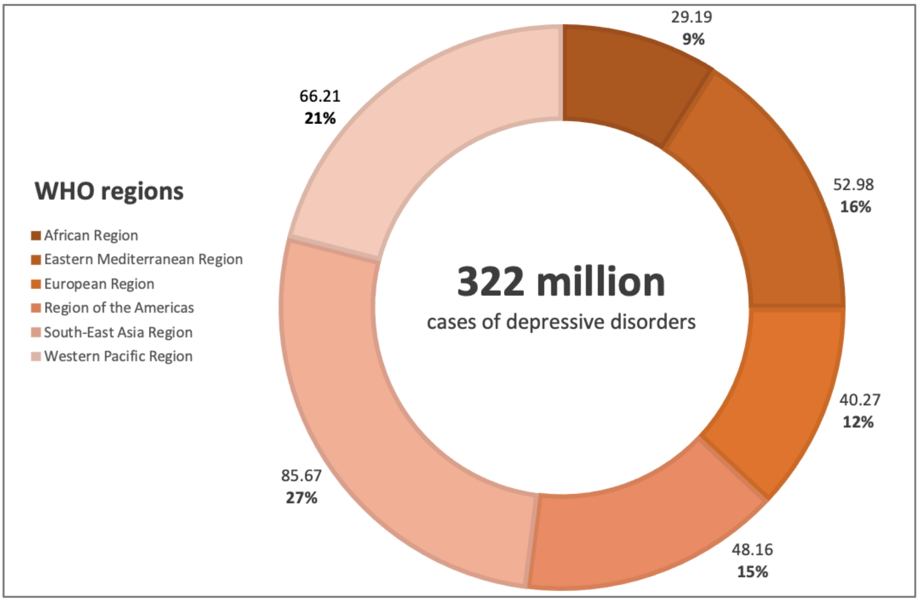
## 1.2.1. Depression prevalence

As it has been demonstrated in several studies, the most common mental disorders are anxiety and depression, the latter being the disorder with high prevalence worldwide in both developed and developing countries and with a major impact on mortality risk by suicide (Lepine, 2011; Mnookin, 2016; Whiteford et al., 2016; WHO, 2017). From 2005 to 2015, the number of people with depression increased by around 18% (Ferrari et al., 2013). In 2015, depressive disorders led to a total of over 50 million YLD (WHO, 2017) being the most disabling disorder (Whiteford et al., 2015; WHO, 2014). More than 80% of depressive conditions take place in low- and middle-income countries and their frequency varies in different WHO regions. Worldwide, depressive disorders are the largest contributor to non-fatal health loss (7.5% of all YLD) (WHO, 2017).

In 2015, more than 300 million people were suffering from depression (equivalent to 4.4% of the world's population), being the main contributor to global disability (WHO, 2017) (see Fig. 2). In other studies, the lifetime prevalence of this disorder was estimated to be between 10% to 15% (Lépine et al., 2011). Depression is linked to most DALYs for both sexes, with a higher percentage in women (Lépine et al., 2011; GBD 2015 DALYs and HALE Collaborators, 2016; Rehm, & Shield, 2019).

**Figure 2**

*Cases of depressive disorder (millions) by WHO Region*




*Note*. Adapted from World Health Organization. (2017). *Depression and other common mental disorders: global health estimates*. Geneva: World Health Organization.

The most recent GBD studies support that depressive disorders are a major cause of non-fatal burden, and in 2017 were the leading cause (GBD 2016 Disease and Injury Incidence and Prevalence Collaborators; GBD 2017 Disease and Injury Incidence and Prevalence Collaborators; 2018). Worldwide, over the 28-year period studied, depressive disorder jointly low back pain and headache disorder were the top causes of YLDs (162 million YLDs in 2017) (GBD 2017 Disease and Injury Incidence and Prevalence Collaborators, 2018).

Depression is the most prevalent mental illness in Spain (Haro et al., 2006). According to the latest national health survey carried out in Spain in 2017, 6.68% of the Spanish population would have diagnosed with depression in the last 12 months (Instituto Nacional de Estadística, 2018) and this means that approximately three million people would be diagnosed with this disease. However, if we take into account that in Spain only 58% of people suffering from depression seek help from health services (Gabilondo et al., 2011), it is possible that this figure does not reflect the total magnitude of the problem.

The amount of people with depression is increasing, especially in lower-income countries, due to more people living to the age when this condition usually (WHO, 2017). The prevalence of this disorder changes depending on age it affects the whole population, from children and adolescents to elderly people, reaching a peak in older adulthood (7.5% in females and 5.5% in males aged 55-74 years old) (WHO, 2017). Depression is an episodic disorder that usually begins early in life, with an average age of onset between the mid to late 20s, however new episodes can appear throughout life. In this kind of patient is typical a pattern of remissions and relapses, with higher risk of recurrence in the cases of early-onset disorder (Lewinsohn et al.; 2000). Many individuals do not recover completely from acute episodes and suffer from a persistent depressive disorder that entails negative consequences on public health worldwide (Gureje, 2011).

Depression is frequently comorbid with other mental disorders (Kessler et al., 2005) such as anxiety (Hyman et al., 2016). This condition increases twice the mortality risk to die prematurely of those who suffer from in comparison to the general population and constitutes an important risk factor for suicide (Isometsa 2014; Hyman et al., 2016; Lépine et al., 2011; Marcus, 2012; WHO, 2017). Likewise, this disorder is frequently associated with cardiovascular, obesity and other general medical disorders such as type 2 diabetes mellitus, coronary artery disease and chronic pain, as well as death by all causes (Lépine et al., 2011; Hyman et al., 2016).

Moreover, depression can be long-lasting or recurrent and impairs cognitive and social functioning. (WHO, 2019a). This disorder has several consequences, both personal and social costs (Lépine et al., 2011; Mather & Locar, 2006; Vigo et al, 2016).

For all of this reasons, the prevention and treatment of depression constitutes a major medical challenge for this century (Lépine et al., 2011). In this sense, the WHO Mental Health Gap Action Program (mhGAP) was created with the intention of promoting, in the countries, services for people with mental, neurological and substance use disorders in the countries, incorporating the care from health workers who are not mental health specialists (WHO, 2018b).

### 1.2.2. Clinical definition of depression

Depressive disorders are characterized by sadness, loss of interest and pleasure, feelings of guilt or low self-worth, disturbed sleep or appetite, feelings of tiredness, and poor concentration (WHO, 2017). The symptoms of depression vary depending on severity of the condition (from mild to severe) and duration (from months to years). These disorders constitute a psychiatric mood disorder that is associated to an individual's inability to cope with stressful events and situations (Hyman et al., 2016). Most people will suffer some form of depression in their lifetime. It is clinically distinguished from normal sadness by its severity, duration and persistence across time (at least two-weeks period) and context (Hyman et al., 2016). These disorders are characterized by a collection of psychological, cognitive, behavioral and physical symptoms that are described in Table 3.

**Table 3**

*Main symptoms that characterize depressive disorders.*

| Depressive disorders symptoms | |
|---|---|
| **Psychological** | - Worthlessness<br>- Feeling critical of oneself<br>- Overwhelming sadness<br>- Anxiety<br>- Feeling "empty" inside<br>- Feelings of guilt<br>- Helplessness<br>- Low self-esteem |

| Depressive disorders symptoms | |
|---|---|
| | - Poor self-image<br>- Preoccupation with death, dying, suicide |
| **Cognitive** | - Thoughts of worthlessness and guilt<br>- Suicidal thoughts<br>- Decreased ability to concentrate<br>- Slowed thinking and speaking-<br>- Ruminations<br>- Trouble recalling details<br>- Increasing preoccupation with depressive feelings<br>- Decreased ability to make decisions |
| **Behavioral** | - Irritability and restlessness<br>- Angry outbursts<br>- Spending increasing amounts of time sleeping<br>- Withdrawing from once-pleasurable activities<br>- Increasing challenges in meeting demands of work, home, social, and scholastic life-<br>- Self-harm<br>- Suicide attempts |
| **Physical** | - Changes in psychomotor movements<br>- Insomnia or hypersomnia<br>- Fatigue and decreased energy<br>- Changes in eating patterns<br>- Aches and pains<br>- Headaches<br>- Digestive problems |

The diagnosis of depression is complex because of the heterogeneous nature of this disease and the diverse manifestation of the symptoms among individuals, which result in a great number of cases that are undetected and untreated, making the prevention, diagnosis, and treatment of the depressive disorders a complicated task (Casano & Fava, 2002; Kraus et. al., 2019; Nambisan, 2015; Nguyen et al., 2017).

Although the biological mechanisms associated to depression are not yet clear enough, several theories related to biochemical, immunological and genetic factors have been suggested. In the last years, the study of the genetic determinants of mental disorders, and in particular depression, has received special attention and, in particular, family studies have provided strong evidence of the influence of genetic factors as a risk for the development of depression. Currently, there are several theories based on the role of genes related to exchange of monoamine neuromediators, chronic stress and the hypothalamic-pituitary-adrenal axis, disturbances of neurogenesis and neuroplasticity, and alterations in the immune system and inflammation processes (Duman & Monteggia, 2006; Jacobs et al., 2000; Shadrina et al., 2018). Furthermore, in relation with genetic factors and their relationship with the establishment of personalized treatments with the currently available psychiatric drugs, pharmacogenetic testing of individuals can contribute for the selection of the best drug treatment for major depressive disorders on the basis of the identification of relevant genetic variants (Pérez et al., 2017; van Westhhenen et al., 2020).

### 1.2.3. Theories of depression

There are diverse psychological theories about depression following psychoanalytic, cognitive, behaviorist, and self-aware schools.

Psychoanalytic theories on depression point out the importance of loss in the onset of depression, above all the loss of love and emotional security. These losses cause a severe and irrational self-criticism of oneself. Freud asserts that depression originates from an excessively severe and dominant super-ego (Freud, 1917).

Beck's (1967) cognitive theory of depression postulates that three mechanisms are responsible for depression that constitute the cognitive triad of negative automatic thinking, the negative self-schemas and the errors of logic or faulty information processing. The first mechanism, the cognitive triad, is made up of negative thoughts about the self, the world, and the future. These negative thoughts are usually automatic and spontaneous and characteristic of people who suffer from depression. People with depression see themselves as helpless, useless, and inadequate. They interpret events in the world in an unrealistic negative way and see the future as totally hopeless. The second mechanism, the negative self-schemas, implies that depressive people possess a depressive schema and, consequently, they see themselves and the world in deeply negative terms. Depressive thinking occurs when these schemas are activated. These negative schemas give rise to the third and final mechanism, errors in logic. When the negative schema is activated, a series of illogical thoughts or cognitive biases control the thinking. In this way, people with negative self-schemas are prone to making logical errors in their thinking, since they selectively focus on certain aspects of a situation ignoring relevant information (McLeod, 2015).

Behavioral theories of depression emphasize the importance of the environment. These theories focus on observable behavior and the conditions through which individuals learn the behavior. In this way, according to classical conditioning, depression is learned by associating certain stimuli with negative emotional states. According to operant conditioning, depression is caused by the removal of positive reinforcement from the environment. Social learning theory holds that depression is learned through observation, imitation, and reinforcement (Lewinsohn, 1974).

The theory of self-awareness for depression postulates that depressed people think a lot about themselves, highlighting the role of self-focused attention (Pyszczynski & Greenberg, 1987). After the loss of a central source of self-worth, people are unable to get out of a self-regulatory cycle focused on recovering what was lost. Consequently, they focus on oneself, magnifying negative emotions and the self-blame, which maintains depressive self-focusing and exacerbates depression (Pyszczynski & Greenberg, 1987). Durkheim (1951) in his theory of the social integration of suicide, argued that the breakdown of equilibrium leads to suicide. The individual has the perception of himself as not integrated into society, a perception of oneself as detached from social life. This social dysregulation influences the depressed person's perceptions of self and causes suicide.

Each of these theories argue a different perspective on the causes and signs of depression. However, many of these theories support the common view that depression influences the way a person perceives themselves and the world around, influencing the person's language and have dramatic effects on the behavior.

### 1.2.4. Language and depression relationship

Language, as a means of communication, constitutes an essential element for providing valuable insights about people's interests, feelings and concerns (Chung & Pennebaker, 2007). It is the most common and reliable way that people use to share with others the thoughts and feelings about their daily lives (Tausczik & Pennebaker, 2010). Tausczik and Pennebaker (2010) pointed out that through language, psychiatrists and psychologists try to understand human beings from the cognitive, personal, clinical and social points of view. Depression modifies the way that people think, feel and communicate and, consequently, the analysis of their language is a useful way to determine how depression affects their feelings and mood. Nevertheless, it is not always easy to understand linguistic behavior and patterns in these patients and, for this reason, it is necessary to carry out further research on the topic.

Previous research on the matter has linked everyday language use with social behavior and personality (Ramirez-Esparza & et al., 2008; Tausczik & Pennebaker, 2010). One approach to analyze people's language is studying the texts they elaborate. In this regard, a well-established link between depression and language use has been observed. The study carried out by Stirman and Pennebaker (2001) examined the language used for the suicidal and non-suicidal poets. The use of first-person singular pronouns was higher in the suicidal poets, with fewer mentions related to the social collective. These findings are consistent with the aforementioned self-awareness and social integration theories.

In this line of research, Rude et al. (2004) considered linguistic patterns of currently depressed, formerly depressed and never depressed students in an essay task. Depressed students utilized significantly more first-person singular pronouns, more negative emotions words and less positive emotion words in comparison to the never depressed students. The use of first-person singular pronouns was more frequent among the depressed people in agreement with other studies (Bucci & Freedman, 1981; Weintraub, 1981). This increased use of first-person singular pronouns demonstrates the attention to self-focus that is consistent with the negative emotional states of depression and the reduced attentional resources, highlighting the psychological distancing to connect with others

(Pennebaker et al, 2003) described by the Beck's cognitive theory of depression (1967) and the self-preoccupation Pyszczynski and Greenberg's theory of depression (1987).

Instead of studying lexical patterns of words, Zinken et al. (2010) focused on the study of the syntactic structures of language, and by analyzing the syntax of texts from depressed patients, they detected improvement of symptoms through the changes in the syntactic structures. Zinken et al. (2010) considered that texts may not differ in the use of words, but they may vary in their syntactic structure and therefore, in the construction of relationships between events. The study concluded that depressed individuals were less likely to use complex syntactic constructs.

In this sense, the diagnosis and severity of depressive disorders is based on different assessment methods such as clinical interviews, skilled observation and self-reports, due to the complexity of these conditions. The careful evaluation of the patients is a prerequisite for an accurate diagnosis (Montgomery, 1980), which means that it should be reproducible, stable, generally accepted and heuristic. For example, one of the goals of rating scales is to increase the forecast accuracy of the clinical prognosis of depression (Lader, 1981) since they contribute to provide systematized insights on the severity of depressive symptoms and to assess the seriousness of the condition, contributing to the decision on treatment options. The periodic assessment can guide treatment and be indicative of progress (Sharp & Lipsky, 2002). The main advantages of depression rating scales are their capacity to evaluate severity and in screening activities (Lee et al., 2010). Each test has its own scoring system, but in all of them higher scores reflect more severe symptoms. In addition, these tests usually have a statistically predetermined cutoff score at which depression symptoms are considered significant and, in some cases, allows us to determine different levels that are linked to symptom severity (Sharp & Lipsky, 2002; Gilbody et al., 2005). Nevertheless, sometimes the lack of consensus when interpreting the rating scales can lead to misdiagnosis of the severity of the patient's depression (Kriston & Wolff, 2011). As it was previously mentioned, language plays a central role in the diagnosis of depression and these clinical instruments and rating scales essentially rely on patient language and its interpretation. Several tests and rating scales for screening depression are shown in Table 4. The review of the language used in these tests, can be used to extract terms related to the expression of symptoms related to depression.

**Table 4**

*Tests and rating scales for screening depression*

| Name of the tests and rating scales for screening depression |
| --- |
| • Beck Depression Inventory (BDI)<br>• Brief Symptom Inventory (BSI)<br>• Carroll Rating Scale for Depression<br>• Center for Epidemiologic Studies Depression Scale (CESD-R)<br>• Clinically Useful Depression Outcome Scale (CUDOS) |

| Name of the tests and rating scales for screening depression |
|---|
| • Goldberg Depression and Anxiety scales (GADS)<br>• Hamilton Rating Scale for Depression (HRSD)<br>• Hospital Anxiety and Depression de Zigmond and Snaith (HAD)<br>• Montgomery-Asberg depression rating scales (MADRS)<br>• The Patient Health Questionnaire (PHQ-9)<br>• Zung Self-Rating Depression Scale (SDS) |

## 1.3. Social media

### 1.3.1. General overview

The Internet revolutionized the way in which we access information, and it had an extraordinary impact on society and culture as well as on different fields such as academia, science, business and all kinds of industries (Castells, 1996). In turn the creation of the World Wide Web (WWW) by Tim Berners-Lee in 1991, which was essential in the creation of the "Information Age" (Castells, 1996), giving the opportunity to millions of people to interact through the Internet. According to the World Wide Web Consortium (W3C), the WWW can be defined as "a global hypertext project that enables people to work together by combining their knowledge in a web of hypertext documents" (W3C, 2020).

From the early 2000s, the Internet has been changing from a consumption information model, the so-called Web 1.0 (Lupiañez et al., 2009), to a user-generated model in which more and more users participate actively in providing and sharing content. This new model is known as the Web 2.0. Tim O'Reilly (2005) described and popularized the term Web 2.0 as:

> Web 2.0 is the network as platform, spanning all connected devices; Web 2.0 applications are those that make the most of the intrinsic advantages of that platform: delivering software as a continually updated from multiple sources, including individual users, while providing their own data and services in a form that allows remixing by other, creating network effects through an architecture of participation, and going beyond the page metaphor of Web 1.o to deliver rich user experiences.

Web 2.0 entails a common vision of its user community, a new perspective in the user's role in the information technologies applications, knowledge and the status of information (Constantinides 2008; Fuchs et al. 2010; O'Reilly 2005; Tredinnick 2006). The evolution of Web 2.0 applications is usually described as "social media". For Gruber (2007), the term social media describes a type of websites and applications characterized for the user participation and user-generated content. For instance, Kaplan and Haenlein (2010, p.61) defined social media as "a group of internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of user generated content”. In this way, the term Web 2.0 is frequently used interchangeably with the term social media, although social media is different from Web 2.0 for other authors (Berthon et al., 2012; Kaplan & Haenlein, 2010; Weinberg & Pehlivan, 2011).

According to Hoffman et al. (2013, p.29) social media is determined by “the set of web-based and mobile tools and applications that allow people to create (consume) content that can be consumed (created) by others and which enables and facilitates connections”. All these different definitions highlight the key principles that characterized the evolution of the web as an open, social and participative environment (Kaplan & Haenlein, 2010; O’Reilly, 2007; Ravenscroft, 2009). This second generation of web technologies open up new types of content generation and sharing, democratizing the exchange of knowledge (Hasan & Pfaff, 2006; Kaplan & Haenlein, 2010; Stenmark, 2008).

Currently, there is a great diversity of social media platforms that offer a multitude of independent and integrated services with common features (Kaplan & Haenlein, 2010; Obar & Wildam, 2015), which are based on the same philosophy of user-generated content that constitutes the essence of Web 2.0 (Kaplan & Haenlein, 2010; Obar & Wildam, 2015). Users of these platforms can access these services via web-based apps on personal computers, laptops and mobile devices (e.g., smartphones and tablets). Individuals, organizations and communities, using their service-specific profiles for the website or app, are able to participate in these interactive platforms, by accessing, sharing, co-creating and modifying user-generated and self-curated content posted online (Boyd & Ellison, 2007; Obar & Wildam, 2015; Treem & Leonardi, 2012). Each social media organization designs and maintains the website or app and facilitates the creation of online social networks and the connection among user's profiles and groups (Boyd & Ellison, 2007; Obar & Wildam, 2015). Table 5 shows the description of the different types of social media platforms.

**Table 5**

*Description of different types of social media platforms*

| Type of social media platform | Description |
|---|---|
| Blogs | It is a chronological list of postings, which can be read and commented by other users. |
| Business networks | Registered users create a personal profile and share different details related to the type and duration of their education, professional experience and expert knowledge. |
| Collaborative projects | These initiatives get together internet users with a common interest or knowledge with the objective to develop, improve, analyze and/or test projects. The results of this collaboration (e.g., programs, findings or games) are distributed as open source and made available to the public free of charge. |

| Type of social media platform | Description |
|---|---|
| Enterprise social networks | These networks are open for registration only to employees of a specific company, ensuring |
| | that their employees know one another and exchange experiences and ideas. |
| Forums | These virtual spaces are discussion platforms where users can ask and/or answer other users' questions and exchange thoughts, opinions or experiences. |
| Microblogs | They are characterized by short entries that can be read by anyone. Postings may include pictures or links. Users can follow other users, institutions, organizations or celebrities. |
| Photo-sharing | These websites allow users to upload, host, manage and share photos. |
| Products/ services review | These reviewing websites provide information about products including customers' reviews. |
| Social bookmarking | These platforms permit users to centralize, save and organize internet bookmarks and share with other users. |
| Social gamming | These online games require social interaction between players. |
| Social networks | These networks connect people that meet each other, share common interests or similar activities. Users have an individual profile to |

| Type of social media platform | Description |
|---|---|
| | share thoughts, pictures or links and to interact with other users. |
| Video sharing | Users may upload and share personal, business or royalty-free videos and to watch them legally. Most websites offer the opportunity to comment on specific videos. |
| Virtual worlds | Users can create a personal avatar and explore the virtual world, participate in its activities or communicate with other avatars. |

*Note.* Adapted from Aichner, T., & Jacob, F. (2015). Measuring the degree of corporate social media use. *International Journal of Market Research*, 57(2), 257-276.

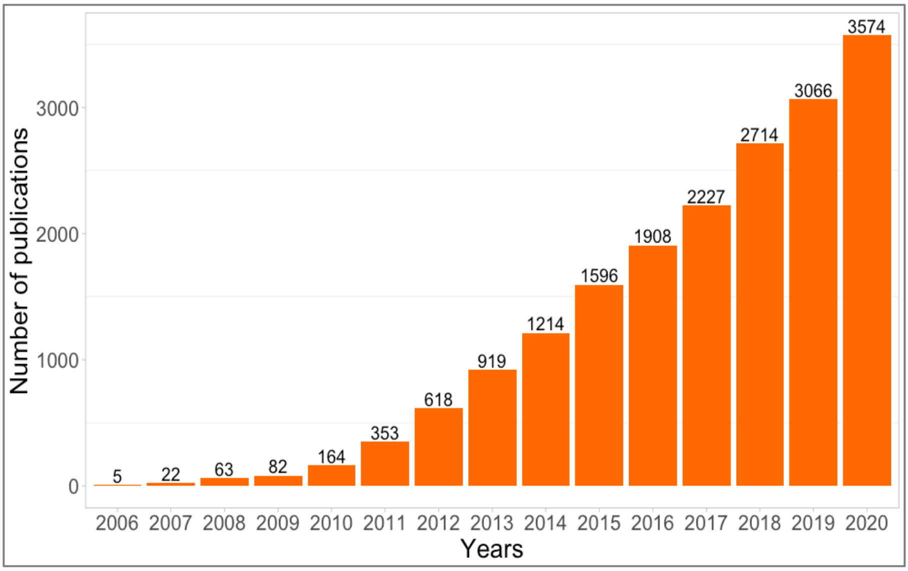
Social media and the way their users participate and interact in these platforms has aroused great interest among scientists, generating a huge number of publications related to this subject. Regarding the biomedical domain, the concept "Social Media" was introduced in 2012 as a main concept in the Medical Subject Headings (MeSH), the most important thesaurus of PubMed, the bibliographic database of reference for scientific publications. Web 2.0, Social Medium and Twitter Messaging are considered entry terms (synonyms, alternate forms or closely related terms) of Social Media in MeSH. According to the MeSH thesaurus, social media are:

Platforms that provide the ability and tools to create and publish information accessed via the INTERNET. Generally, these platforms have three characteristics with content user generated, high degree of interaction between creator and viewer, and easily integrated with other sites (National Library of Medicine, 2012).

Since its first mention in 2005, the interest in this subject has been increasing every year from the scientific point of view based on the number of publications included in PubMed as is shown in Figure 3.

**Figure 3**

*Number of publications in PubMed about Social Media in the last 15 years*



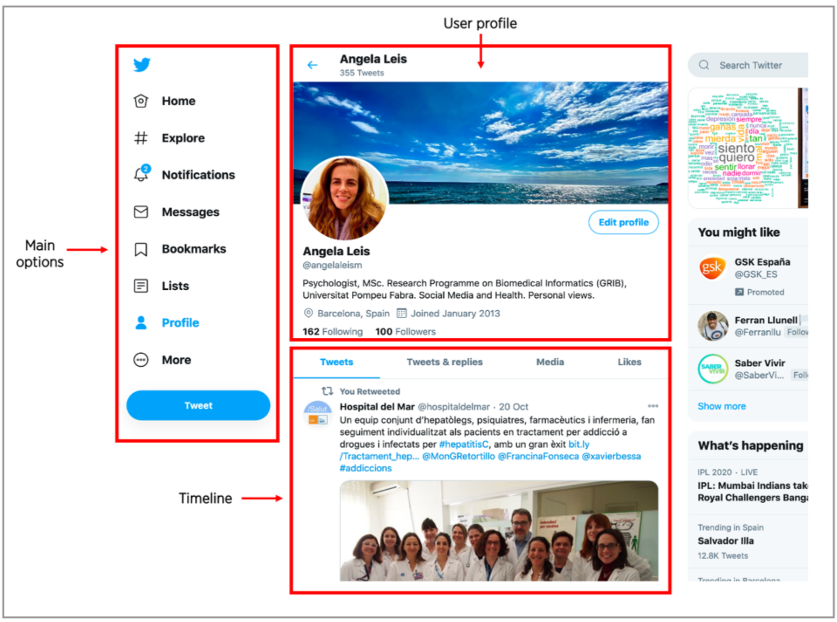
## 1.3.2. Twitter

Twitter is a free microblogging and social networking service created by Jack Dorsey that was launched in 2006. It is one of the most important social media platforms in terms of number of users, totalizing more than 330 million active users worldwide (Clement, 2019).

Each Twitter user can create one or more accounts including basic personal information although this is not necessary. It is possible to customize the account and to decide if it should be protected or open to anyone (Meskó, 2013). In Figure 4, the homepage of a user is shown.

**Figure 4**

*Twitter account homepage*



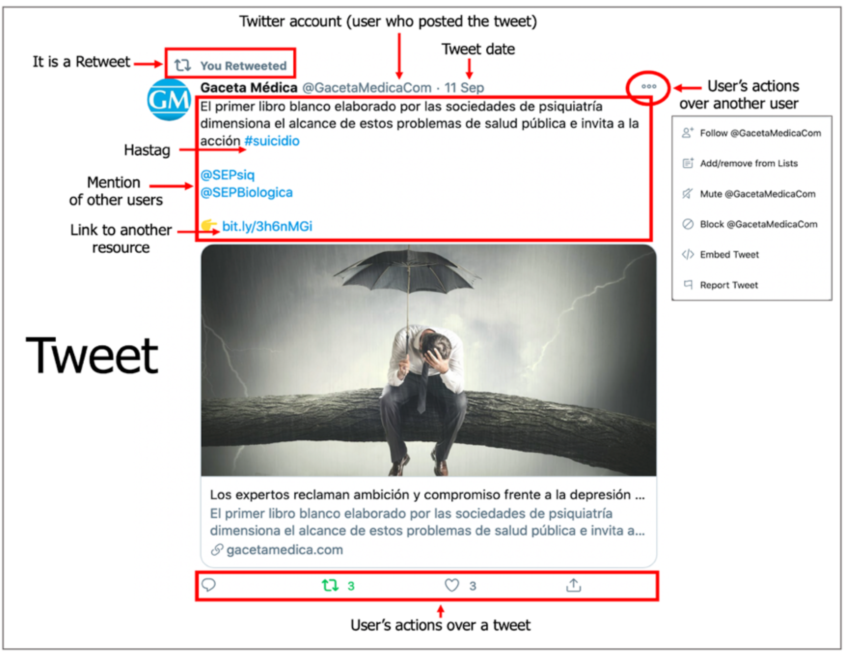
A tweet is an online posting created by a Twitter user. You can publish a tweet using a computer, a tablet or a mobile phone. The main definitions related to a tweet are:

- Tweets: 280 characters (text, links, multimedia)
- Retweets: RT (MT: RT modified)
- Answering someone: @angelaleism
- Direct message: D (only if it is a follower)
- Bookmarks: bookmarking and sharing tweets
- Hashtag ("#"): expressions used to tag messages with a #topic
- Following (following someone)
- Followers (someone's followers

In Figure 5, a tweet and its main elements are shown.

**Figure 5**

*Main elements of a posted tweet*



Since November 2017, the maximum number of characters of a tweet has been increased from 140 to 280. In almost 90% of users' accounts, the tweets are freely available and provide a huge amount of data that can collected in real-time (Alvaro et al., 2015; Audeh et al., 2020; Eysenbach, 2009; Nikfarjam et al., 2015; Pierce et al., 2017; Sarker et al., 2015;). The users from Spanish-speaking countries are among the most active on Twitter in the world (Clement, 2019).

### 1.3.3. Digital phenotype

Social media platforms were not set up with health-related purposes in mind (Salathé, 2016; Sarker et al., 2015), but millions of people publicly share personal health information on them (Adrover et al., 2015; Nikfarjam et al., 2015). Consequently, social media represents an important source of health information, which is more easily and broadly available than other sources of health information, being unsolicited, spontaneous and permanently updated (Salathé, 2016; Sarker et al., 2015).

The widespread access to digital technologies such as smartphones, mobile applications, wearable devices and social media, have changed the way people communicate and interact with each other and with the world around them (Naslund, Gonsalves, Gruebner et al., 2019). In 2020, over 3.6 billion people were using social media worldwide and it is expected that the number of users will increase to more than 4.4 billion in 2025 (Clement, 2020).

In this context, infodemiology approaches have been developed and applied to better understand the dynamics of these platforms when used as a health information source (Eysenbach, 2009; Lardon et al., 2018; Salathé, 2018). The infodemiology term was established by Eysenbach (2009) in the following way "infodemiology can be defined as the science of distribution and determinants of information in an electronic medium, specifically the Internet, or in a population, with the ultimate aim to inform public health and public policy".

The use of infodemiology for surveillance purposes, the so-called infoveillance (Eysenbach, 2009), focus on public health related concerns, and use online information for monitoring them. In this sense, social media can be used as an additional tool for health monitoring and surveillance. Considering that social media users share relevant health-related information, such as experiences with prescribed drugs (Freifeld et al., 2014), cancer patients' sentiments (Crannell et al., 2016), opinions on vaccines (Radzikowski et al., 2016; Surian et al., 2016), diabetes sentiment (Gabarron et al., 2019), rare diseases or healthy eating seeking support (Leis et al., 2013; Subirats et al., 2018), online conversations on epidemic outbreaks such as COVID-19 or Zika (Chen et al., 2020; Finch, et al., 2016; Stefanidis et al., 2016), this information can be used as a complementary source for decision-making. Online sources of big data provide new possibilities to study users' behavior (Nikfarjam et al, 2015), detect patterns in human behavior and illness trajectories (Naslund, Gonsalves, Gruebner et al., 2019), health dynamics in populations (Salathé, 2018), monitor disease risk (Llieva & McPhearson, 2018) and to inform providers about the development of targeted interventions (Insel, 2018). The use of data generated outside the formal channels used by the public health systems constitute an emerging field, also called digital epidemiology (Salathé, 2018).

The human interaction with social media tools contributes to build the so-called digital phenotype, reshaping disease expression in terms of the lived experience of individuals and detecting early onset of several conditions (Jain et al., 2015). In the same way, the daily interactions of people connected to the Internet generates a trail of data that has been defined by different authors as digital footprints (Bidargaddi et al., 2017; Gittelman et al., 2015; Kosinski et al., 2013; Youyou et al., 2015), expanding the previous definition of digital footprints to include the data generated during the interaction with smartphones and other devices or gadgets. These digital footprints allow us to obtain data in real time from different personal contexts (social, occupational and domestic) of daily life, which can influence the person's mental state (Bidargaddi et al., 2017). These data characterize the behavior of individuals and their environment, reflecting their longitudinal changes, which linked to mental health data, can contribute to a better understanding of the etiology of psychiatric problems (Bidargaddi et al., 2017).

These digital footprints can be an important information resource for epidemiological research of psychiatric disorders, contributing to the refinement of the diagnostic criteria or to monitor the response to treatment. Some studies have demonstrated the suitability of mobile phone data to obtain various behavioral indicators of mental health (Glenn & Monteith, 2014), their contribution to the characterization of the longitudinal phenotype in psychiatry (Wenzel et al., 2016), and their potential for collecting such data from psychiatric patients (Ben-Zeev et al., 2015; Saeb, 2015; Schwartz, 2016).

Finally, it is worth mentioning that users are surprisingly willing to share their data for research purposes and for the benefit of other patients (Ennis et al., 2012). Consequently, there is a need to develop strategies that offer people better control of their diverse digital footprints giving the opportunity of deciding what kind of information they want to share in order to balance the risk and benefits of using such data.

### 1.3.4. Social media and depression

The social nature of social media platforms, such as Facebook, Twitter, Reddit or Instagram among many others, allows us to detect social patterns, thereby revealing key aspects of mental disorders (Coppersmith, Drezde & Harman, 2014). Nowadays these platforms have become an important source of health-related information, including that related to mental disorders. The huge amount of health data present in social media can be monitored and analyzed by using advanced tools based on big data analytics such as text mining, natural language processing (NLP) and machine learning technologies. Text mining is “the task of extracting meaningful information from text”, and NLP is a “subfield of computer science, artificial intelligence and linguistics which aims at understanding of natural language using computers” (Allahyari et al., 2017). The use of these methodologies has shown to be effective in supporting and automating the identification of early signs of mental illness in their users by analyzing the content shared in the Web (Conway & O’Connor, 2016; Paul & Drezde, 2011; Park et al., 2012). These methods for studying mental health, an emerging research area, bring new opportunities to develop research and interventions in mental health. Furthermore, these methods can expand our understanding of mental disorders with the aim of advancing in the implementation of early and effective interventions, having a particularly beneficial impact in lower income healthcare services (Naslund, Gonsalves, Gruebener et al., 2019).

The analysis of the messages posted on social media platforms may provide information about many personality traits, lifestyles, and psychological disorders (Paul & Dredze, 2011; Prieto et al., 2014; Thackeray et al., 2013). By analyzing huge amounts of text, researchers can link everyday language use with social behavior and personality (Tausczik & Pennebaker, 2010; Ramirez-Esparza & et al., 2008). The perceived anonymity of social media encourages its users to be more willing to self-report health information, such as details of their mental disorders and treatments received. In addition, such media are seen as a way to communicate and receive support from others with similar experiences, avoiding in some cases the isolation and social stigma of these conditions (Berry et al., 2017; Coppersmith et al., 2015; Ferrari et al, 2013; Nguyen et al., 2015; Pavalanathan & De Choudhury, 2015; Prieto et al., 2014).

There are many studies, usually focused on messages written in English, which have used data mining and machine learning techniques on social media platforms to automatically identify people with mental health problems, such as posttraumatic stress disorder, schizophrenia, eating disorders or suicidal ideation among others (Arseniev-Koehler et al., 2016; Birnbaum et al., 2017; Conway & O'Connor, 2016; Coppersmith, Harman & Drezde, 2014; Hswen et al., 2018; Wang et al., 2018; Wilson et al., 2014) and to detect seasonal patterns of searches performed on the Internet about psychiatric disorders (Soreni et al., 2019).

Considering that depression is the most prevalent mental disorder, there are several studies that analyze the different forms of expression and manifestations of depressive conditions on social media platforms. These studies have found that people with depression often use social media to talk about their illness and drugs they are taking, share information and experiences, seek social support and advice trying to reduce their social isolation, and manage their mental illness (Berry et al., 2017; Cavazos-Rehg et al., 2016; Coppersmith et al., 2015; De Choudhury, Gamon, Counts et al., 2013; Naslund et al., 2019; Nguyen et al., 2015; Wilson & Valstar, 2014).

Several studies were carried out mining Twitter about depressive symptoms with the aim of detecting and monitoring depression. In these studies diverse features of post messages were analysed, such as the number and frequency of tweets, distribution throughout the day and night hours and their seasonal character, frequencies of words of different grammatical categories, and emotions associated with the words used (Cavazos et al., 2016; De Choudhury, Counts, Horvitz, 2013; De Choudhury, Gamon, Counts et al., 2013; Mowery et al. 2017; Nambisan et al., 2015; Park et al., 2012;  Reece et al.; 2017; Wilson & Valstar, 2014).

Other studies were focused on the detection of markers of depression in posts on Instagram, using images, text and behavior features (Reece & Danforth, 2017; Chiu et al., 2020; Reece & Danforth, 2017), observing that the community-generated data can constitute a complementary data to identify depression among social media users (Ricard et al., 2018). In addition, the analysis of Internet forums showed that absolutist words tracked the severity of affective disorders more faithfully than negative emotion words (Al-Mosaiwi & Johnstone, 2018). Also, several studies showed the utility of analyzing social media platforms to detect depression, such as the analysis of language used on Facebook to predict the diagnosis of depression stated in electronic health records (Eichstaedt et al., 2018), the detection and prediction onset of post-partum depression on Facebook as well (De Choudhury, Counts, Horvitz et al, 2014) or the detection of depression related posts in Reddit (Tadesse et al., 2019).

Likewise, mining drug-related information from Twitter has been applied and several studies have shown that social media can be used as a complementary source for pharmacovigilance and monitoring (Carbonell et al., 2015; Freifeld et al., 2014; O'Connor et al., 2014; Salathé, 2016). This approach can be useful to identify users' mentions of drugs intake on social media, to obtain information on how patients respond to the pharmacological treatments they are receiving (Kiritchenko et al., 2017; Klein et al., 2017; Mahata et al., 2018), or to analyze users' tweets regarding adverse events associated to a particular treatment (Adrover et al., 2015; Nikfarjam et al., 2019; Pierce et al., 2017; Segura-Bedmar et al., 2015).

According to some studies, it is common that patients suffering from depression do not maintain the duration of antidepressant treatment that is clinically recommended (Royal College of Psychiatrists, 2019; National Institute for Health and Care Excellence (NICE), 2018; De Choudhury & De, 2014). Taking into account that the importance of adherence to treatment is essential for disease remission (Anghelescu et al., 2006; De las Cuevas et al., 2014; Mitchell, 2016; Martin-Vazquez, 2016), the use of social media-based approaches to monitor patients constitutes an attractive opportunity for the detection of changes in symptoms when patients are taking medications, which not only provide interesting insights for monitoring pharmacological treatments, but also for controlling the evolution of the disease and the treatment adherence.

This emerging area of research can help healthcare professionals and institutions in the decision-making processes to ensure better management of patients suffering from depression.

# 2. OBJECTIVES

*"Getting information off the Internet is like*
*taking a drink from a fire hydrant"*
Mitch Kapor

Mental disorders in general, and depression in particular, constitute a public health concern worldwide. The availability of innovative instruments and tools for the diagnosis and monitoring of these diseases is foremost.

The detection and interpretation of depression-related behavioral and linguistic patterns in messages on social media environments such as Twitter have the potential to constitute a new tool of the detection and follow-up of depressive patients, including their pharmacological treatment.

There are no studies about depression that analyze messages on social media written in Spanish. Taking into account that Spanish speaking countries, are among the most active Twitter users in the world, the analysis of this information using big data analytics can contribute to better understand and monitor depressive users.

The general objective of this thesis is to study the usefulness of social media for detecting and characterizing behavioral and linguistic patterns in messages on social media in Spanish language and linking them to the expression of depression and to effect of antidepressant medication.

The specific objectives are the following:

1. To identify the linguistic features of tweets and the behavioral patterns of the corresponding Twitter users that could suggest signs of depression.

2. To determine the usefulness of text mining and natural language processing tools for the analysis of tweets in Spanish language.

3. To create a comprehensive collection of Spanish words commonly used by patients suffering from depression, which will be used for detecting depression in Twitter users.

4. To detect the existence of changes in behavioral and linguistic features of Twitter users associated with the mention of antidepressant medication.

# 3. DESIGNING A STUDY ON TWITTER

*"You affect the world by what you browse"*

Tim Bernes-Lee

The definition and implementation of proper computational approaches to consistently access and retrieve content of interest from Twitter (i.e., tweets with particular characteristics) constitutes an essential step to be carried out before any further analysis of the data. Twitter provides users with an Application Programming Interface (API), described in more details hereinafter, useful to enable programmatic data retrieval from its platform.

This section describes: (i) the data structure of a tweet (i.e., the pieces of information provided by Twitter in order to describe a tweet, including its textual content, the ID of the user publishing the tweet, the publication time, etc.); (ii) the API of Twitter and the way such API has been exploited to crawl tweets; (iii) the approach followed to properly analyze the content of tweets with the aim of extracting the linguistic and semantic features considered in this dissertation.

## 3.1. Data structure of a tweet

Tweets are the basic blocks of information used to share content on Twitter. Tweets are also known as "status updates". Each tweet is characterized by a long list of attributes or fields, describing several pieces of information useful to characterize it, including an unambiguous identifier assigned by Twitter (*id* field), the date the tweet has been posted (*created at* field), the text of the tweet (*text* field) and a set of metadata describing the author of the tweet (*user* field) (Twitter, 2020b). The contents of tweets are distributed by

Tweeter by relying on the JavaScript Object Notation (JSON) data-interchange format. As an example, Figure 6 shows a tweet as rendered by the Twitter Web application, while Figure 7 provides the JSON-based serialization of the contents of the same tweet.

**Figure 6**

*A tweet in a human readable format*


*Note.* Obtained from Kevin Pho [@kevinmd]. Physician burnout starts in medical school. But it doesn't have to [Tweet].

The rendered JSON format of the previous tweet will be a mix of different attributes or fields. Below, an example that includes some of the most fundamental attributes and child objects, which are represented here with the { } notation, corresponding to the information contained in the previous human readable tweet (Figure 6).

**Figure 7**

*JSON-based serialization of tweet content of Figure 6*


```
{
"created_at": "Sat Sep 05 15:02:52 +0000 2015",
"id": 640797582453788672,
"id_str": "640178287629631489",
"text": "RT @kevinmd: Physician burnout starts in medical
school.  But it doesn\u2019t have to.
http://t.co/KMIbpSWXd5 http://t.co/IuRuWLSJ4R",
"user": { },
"entities": { }
}
```


Twitter defines a data format useful to exploit the content of each tweet. Below in the Table 6, the information concerning the most relevant root-level attributes used in the analysis performed in this thesis are shown, including the name of the variable or attribute, the type of variable and its description:

**Table 6**

*Most relevant attributes used in the analysis performed*

| Attribute | Type | Description |
|---|---|---|
| created_at | String | UTC time when this Tweet was created. Example: "created_at": "Wed Oct 10 20:19:24 +0000 2018" |
| id | Int64 | Integer representation of the unique identifier for this Tweet. This number is greater than 53 bits and some programming languages may have difficulty/silent defects in interpreting it. Using a signed 64 bit integer for storing this identifier is safe. Use id_str to fetch the identifier to be safe. See Twitter IDs for more information. Example: "id":1050118621198921728 |
| id_str | String | String representation of the unique identifier for this Tweet. Implementations should use this rather than the large integer in id. Example: "id_str":"1050118621198921728" |
| text | String | Actual UTF-8 text of the status update. See *twitter-text* for details on what characters are currently considered valid. Example: "text": "To make room for more expression, we will now count all emojis as equal-including hose with gender and skin t… https://t.co/MkGjXf9aXm" |

| Attribute | Type | Description |
|---|---|---|
| user | User object | User who posted this Tweet. See User data dictionary for complete list of attributes.<br>Example<br>highlighting<br>select attributes:<br>{ "user": {<br>"id": 6253282,<br>"id_str": "6253282",<br>"name": "Twitter API",<br>"screen_name": "TwitterAPI",<br>"location": "San Francisco, CA",<br>"url": "https://developer.twitter.com",<br>"description": "The Real Twitter API. Tweets about API changes, service issues and our Developer<br>Platform. Don't get an answer? It's on my website.",<br>"verified": true,<br>"followers_count": 6129794,<br>"friends_count": 12,<br>"listed_count": 12899,<br>"favourites_count": 31,<br>"statuses_count": 3658,<br>"created_at": "Wed May 23 06:01:13 +0000 2007",<br>"utc_offset": null,<br>"time_zone": null,<br>"geo_enabled": false,<br>"lang": "en",<br>"contributors_enabled": false,<br>"is_translator": false,<br>"profile_background_color": "null",<br>"profile_background_image_url": "null",<br>"profile_background_image_url_https": "null",<br>"profile_link_color": "null", |

| Attribute | Type | Description |
|---|---|---|
| | | "profile_sidebar_border_color": "null",<br>"profile_sidebar_fill_color": "null",<br>"profile_text_color": "null",<br>"profile_use_background_image": null,<br>"profile_image_url": "null",<br>"profile_image_url_https":<br>"https://pbs.twimg.com/profile_images/<br>942858479592554497/<br>BbazLO9L_normal.jpg",<br>"profile_banner_url":<br>"https://pbs.twimg.com/profile_banners/<br>6253282/1497491515",<br>"default_profile": false,<br>"default_profile_image": false,<br>"following": null,<br>"follow_request_sent": null,<br>"notifications": null<br>}<br>} |
| entities | Entities | Entities which have been parsed out of the tex of the Tweet. See Entities in Twitter Objects f additional information. Example:<br>"entities":<br>{<br>"hashtags":[],<br>"urls":[],<br>"user_mentions":[],<br>"media":[],<br>"symbols":[]<br>"polls":[]<br>} |
| retweeted | Boolean | Indicates whether this Tweet has been Retweeted by the authenticating user. Example: "retweeted":false |

| Attribute | Type | Description |
|---|---|---|
| lang | String | *Nullable.* When present, indicates a BCP 47 language identifier corresponding to the machine-detected language of the tweet text, or "und" if no language could be detected. Example: "lang":"en" |

*Note.* Obtained from Twitter Developer. Getting started guide.

## 3.2. Twitter Application Programming Interface (API)

A properly authorized Twitter developer account is needed in order to get access to the Twitter API and retrieve data from Twitter. This account allows users to access the different services (i.e., API endpoints) that Twitter offers to obtain tweets. Below we provide an overview of the four fundamental steps required to activate a Twitter developer account:

**1) The user has to apply and receive approval for a developer account.**

To make any request to the Twitter API, the user must first apply for a developer account, and have the use case approved. Once approved, the user can create a project and connect an associated Developer App which will provide a set of credentials that the user will use to authenticate all requests to the API.

**2) The user has to save the App's Key and Tokens and keep them secure.**

Within the developer App, the user will be provided of a set of API Keys (also known as Consumer Keys). The user will also have the possibility to generate a set of Access Tokens (the security credentials for a login session that identify the user), which can be used to make requests on behalf of their personal Twitter account.

**3) Set up the user's access.**

At this point in time, there are a few different access tiers across two different versions of the Twitter API endpoints that require different provisioning and authentication methods.

**4) The user can make the first request.**

Once the user has access, she is ready to get started using the Twitter API. Twitter has quick start guides for many of its endpoints, including several useful tools, libraries, and tutorials.

Once a developer account is created, several Twitter API clients spanning over a varied set of programming languages are available to access Twitter accounts and retrieve tweets.

## 3.3. Twitter crawler

A crawler is a computer program that explores Web content in order to gather information of interest. In this thesis a Java-based tool developed by Dr. Francesco Ronzano of the Research Programme on Biomedical Informatics (GRIB, IMIM-UPF) was used in order to crawl Twitter for obtaining the tweets that were subsequently analyzed. The different necessary steps for configuring the Twitter crawler are described in the following website: https://github.com/fra82/twitter-crawler.

We decided to use the Twitter API, filter tweets by Spanish language and use the set of keywords described in the Results (sections 4.2 and 4.3) in order to select tweets of interest. The set of keywords used to filter content from Twitter stream is usually specific to the topic of each study. It is worth mentioning that Twitter applies an algorithm using that makes use of diverse features of the tweets, such as linguistic content and words used or users' profiles, to assign a language to each tweet. In our case, tweets were not filtered by geographical information due to the slight percentage of users providing this information.

- Tweets and retweets were collected, even though retweets were excluded from all the analyses except the analysis that required to know the total number of tweets and retweets posted per hour and the number of tweets and retweets issued throughout the week.

- The fields of tweets used for the analysis were: user id, text of the tweet, tweet id, date of the tweet and time zone. Time zones were normalized to Greenwich Mean Time (GMT) using a program in order to make time across distinct time-zones comparable.

## 3.4. Preprocessing and normalization of text

Once the tweets of interest were collected, it is necessary to process them applying several procedures:

- A list of 210 stopwords in Spanish have been removed from the text of each tweet. Stopwords are the most commonly used words in a language; they are usually filtered out when textual data is analyzed. Indeed, stopwords are words that do not provide any useful information because they do not have a defined meaning, or they are too frequent. Some examples of these words in Spanish are: *un, una, también, en, por, encima, arriba, uso, lo, la*, and so on. The complete list of these words is provided in Appendix 4. The images, links and emojis included in each tweet were not analyzed due to the extremely low percentage of tweets including them. All the words included in the tweets were analyzed by applying the same approach, treating the hashtags as words by removing the hash char ('#'). Hashtags are words preceded by a hash sign (#) that are used in microblogging services such as Twitter.

The textual content of each tweet was analyzed by means of the following sequence of steps:

- Tokenization of words (word and punctuation splitting) and punctuation removal, thus creating a list of words. After that, the words were grouped into sentences again. The tokenization process was performed by means of a custom Twitter tokenizer included in the Natural Language Toolkit (NLTK Project, 2020).
- A Part of Speech (POS) was assigned to each token (words). In this process the words are classified according to their grammatical category and tagged (part-of-speech tags).

  The Part-of-Speech (POS) tagging process were performed by means of the Freeling Natural Language Processing (FreeLing) tool in order to analyze the usage patterns of grammatical categories (e.g., adjectives, nouns, or pronouns) in the text of tweets (Padró & Stanilovsky, 2012). FreeLing is an open-source language analysis tool, which consists of a C++ library providing language analysis functionalities for different languages including Spanish and provides the output in the desired format (XML, JSON, CoNLL). These tools are developed and maintained by the TALP Research Center, in Universitat Politècnica de Catalunya. There are other natural language processing (NLP) tools tailored to analyze Spanish texts, such as the CoreNLP created by the Natural Language Processing Group at Stanford University. Nevertheless, we used FreeLing because we consider that is a robust tool for NLP (Natural Language Processing) in Spanish.

Other linguistic analyses that were carried out were:

- Identification of negations performed by relying on a custom list of Spanish negation expressions, such as *nada* (nothing), *nadie* (nobody), *no* (no), *nunca* (never), and alike: we considered exclusively negations that are expressed using simple words, referred as to sentential negations.
- Identification of occurrences of positive and negative words inside the text of each tweet by means of two Spanish polarity lexicons:
    1. The Spanish Sentiment Lexicon (Pérez-Rosas et al., 2012). This lexicon consists of 1,347 words that are classified in two categories, positive and negative polarity words.
    2. The Spanish SentiCon Lexicon (Cruz et al., 2014). This lexicon consists of 11,542 words with a score of polarity between -1.0 (negative) and 1.0 (positive).

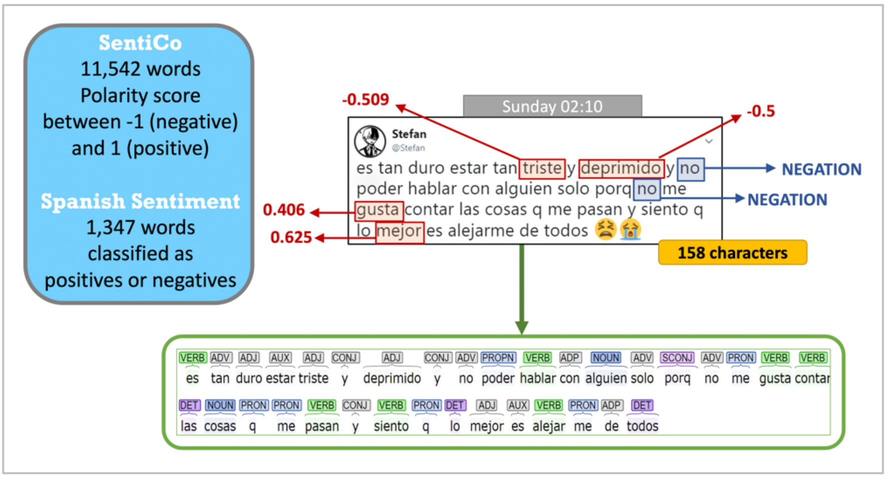
  We exploited these different lexicons to consider and compare two approaches of modeling polarity in Spanish texts, thus reducing any language modeling bias that the use of a single language resource could introduce. In Figure 8, an example of polarity analysis of a tweet is show.

  Some examples of words relate to positive and negative polarity are:

- Positive polarity: adorar, digno, educado, conocido, energizar.
- Negative polarity: erróneo, gamberro, humillación, fallo, oprimir.

**Figure 8**

*An example of polarity analysis of a tweet*



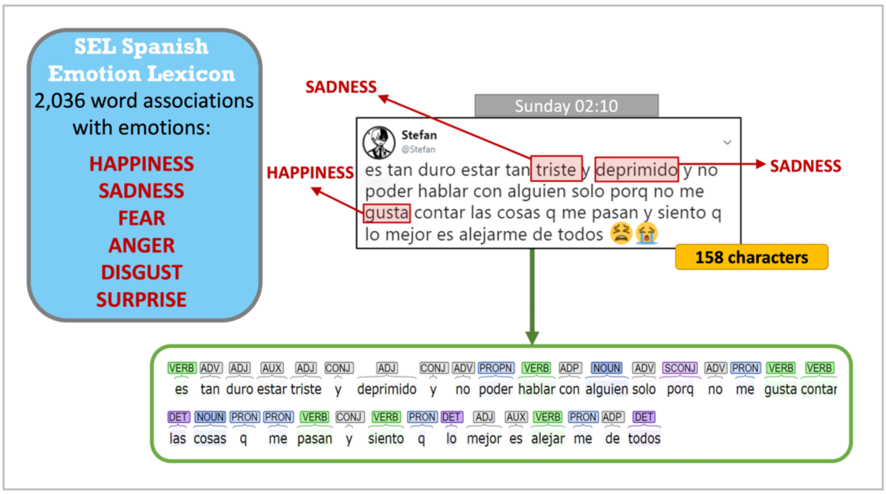

- Identification of words and expressions associated with the basic emotions (Ekman et al., 1987) by using the Spanish Emotion Lexicon (Sidorov et al., 2012), that contains 2,036 words. Such emotions are: *alegría* (happiness), *enojo* (anger), *miedo* (fear), *repulsión* (disgust), *sorpresa* (surprise), and *tristeza* (sadness). The emotions were counted taking into account the total number of words of each emotion in all the tweets (bag of words). An example of emotion analysis of a tweet is shown in Figure 9.

Some examples of Spanish words regarding de different emotions are:

- Anger: *desprecio (scorn)*, *enfado (annoyance)*.
- Fear: *angustia (anguish)*, *impotente (helpless)*.
- Disgust: *asco (disgust)*, *desagradable (unpleasant)*.
- Happiness: *amor (love)*, *amistad (friendship)*.
- Sadness: *agobiado (overwhelmed)*, *culpable (guilty)*.
- Surprise: *despistado (absent)*, *susto (fright)*.

**Figure 9**

*An example of emotion analysis of a tweet*




All the tools and aforementioned resources are free and publicly available.

# 4. RESULTS

*"There are three kinds of lies:*

*lies, dammed lies, and statistics"*

Marc Twain

## 4.1. Clinical-based and expert selection of terms related to depression for Twitter streaming and language analysis

People use language to express their thoughts and feelings, unveiling important aspects of their psychological traits and social interactions. Although there are several studies describing methodologies to create a collection of words in English related to depression, in most of them the selection of words is not expert based. The objective of this study is twofold: firstly, to introduce a comprehensive collection of Spanish words commonly used by patients suffering from depression, and secondly, to study the usefulness of this collection of words in identifying social media posts that could be indicative of patients suffering from depression. The level of agreement among medical doctors to determine the best words that should be used to select tweets related to depression was low. This finding may be due to the complexity of depression and the diversity when describing the illness. The words supposedly more linked to depression are very common words used in other contexts, and less specific for detecting depressive users.

Leis, A., Mayer, M. A., Ronzano, F., Torrens, M., Castillo, C., Furlong, L. I., & Sanz, F. (2020). Clinical-Based and Expert Selection of Terms Related to Depression for Twitter Streaming and Language Analysis. *Studies in health technology and informatics, 270*, 921.925. https://doi.org/10.3233/shti200296

Digital Personalized Health and Medicine
L.B. Pape-Haugaard et al. (Eds.)


doi:10.3233/SHTI200296

# Clinical-Based and Expert Selection of Terms Related to Depression for Twitter Streaming and Language Analysis

Angela LEIS[a], Miguel-Angel MAYER[a], Francesco RONZANO[a], Marta TORRENS[b,c], Claudio CASTILLO[b], Laura I. FURLONG[a] and Ferran SANZ[a,1]
[a] *Research Programme on Biomedical Informatics (GRIB), Hospital del Mar Medical Research Institute (IMIM) and Universitat Pompeu Fabra, Barcelona, Spain*
[b] *Institute of Neuropsychiatry and Addictions, IMIM, Barcelona, Spain*
[c]*Psychiatric Department of Universitat Autonoma de Barcelona, Spain*

**Abstract.** People use language to express their thoughts and feelings, unveiling important aspects of their psychological traits and social interactions. Although there are several studies describing methodologies to create a collection of words in English related to depression and other conditions, in most of them the selection of words is not clinical or expert based. The objective of this study is twofold: firstly, to introduce a comprehensive collection of Spanish words commonly used by patients suffering from depression, which will be available as a free open source for research purposes (GitHub), and secondly, to study the usefulness of this collection of words in identifying social media posts that could be indicative of patients suffering from depression. The level of agreement among medical doctors to determine the best words that should be used to select tweets related to depression was low. This finding may be due to the complexity of depression and the extraordinary diversity in the way people express themselves when describing their illness. It is critical to perform a thorough analysis of the specific language used in each condition, before deciding the best words to be used for filtering the tweets in each disease. As our study shows, the words supposedly more linked to depression are very common words used in other contexts, and consequently less specific for detecting depressive users. In addition, grammatical gender forms should be considered when analysing some languages such as Spanish.

**Keywords.** Depression, social media, surveys and questionnaires, terminology

## 1. Introduction

People use words to reflect their thoughts and feelings, revealing a huge amount of information about their personality and social interactions, as well as different psychological traits [1]. Language is the medium by which mental health professionals attempt to understand human beings and mental disorders. Several studies have found that linguistic styles are indicative of depressed mood [2]. Another interesting point is that people's language is stable over time and consistent across subjects or context; consequently, it can be used as a tool to measure differences among individuals [3].

[1] Corresponding Author: E-mail: ferran.sanz@upf.edu.

Several studies used social media platforms such as Twitter to analyze psychiatric symptoms and diseases, including depression [4,5]. Although these studies describe different methodologies to create a collection of words in English related to depression and other mental disorders, in the majority of them the selection of words were not extensive or clinical and expert-based [6,7]. In addition, Spanish speaking countries, such as Spain and Mexico, are among the ten countries with most Twitter users worldwide, with more than 6 million and 7 million users, respectively [8]. As far as we know, there are no studies focused on the creation of a list of words of depression in Spanish. The selection of the best terminology and keywords in any information system, including electronic health records, bibliographic databases or social media platforms, is critical for their usefulness in management and scientific research.

The objective of this study is twofold: first, to create a comprehensive collection of Spanish words commonly used by patients suffering from depression, which will be used for streaming Twitter and that is available as a free open source for research purposes in GitHub. Second, we determine the usefulness of the different words in identifying social media posts potentially related to depression.

## 2. Methods

The methodological approach consisted of three phases. The first phase consisted of the review of the most common tests and rating scales for assessing depression (using their Spanish version) in order to extract terms related to the expression of the depression symptoms. The different tests and rating scales analyzed are shown in table 1. The review was carried out by a psychologist and a family physician with experience in the clinical characteristics of depression, who reached an agreement on the words representative of the language used by depressive patients. The Diagnostic and Statistical Manual of Mental Disorders, fifth edition (DSM-5) [9] was reviewed to complete the list. In this way, a list of 255 words was created.

**Table 1**. Tests and rating scales used for selecting words related to the expression of depression by patients (the Spanish versions were used)

| Name of the tests and rating scales |
|---|
| Beck Depression Inventory (BDI) |
| Brief Symptom Inventory (BSI) |
| Carroll Rating Scale for Depression |
| Center for Epidemiologic Studies Depression Scale (CESD-R) |
| Clinically Useful Depression Outcome Scale (CUDOS) |
| Goldberg Depression and Anxiety scales (GADS) |
| Hamilton Rating Scale for Depression (HRSD) |
| Hospital Anxiety and Depression de Zigmond and Snaith (HAD) |
| Montgomery-Asberg depression rating scales (MADRS) |
| The Patient Health Questionnaire (PHQ-9) |
| Zung Self-Rating Depression Scale(SDS) |

In the second phase, a questionnaire was created including the aforementioned list of words. The questionnaire was sent in December 2017 via email to 50 psychiatrists from the Institute of Neuropsychiatry and Addiction (INAD) of Parc Salut Mar in Barcelona and 5 family physicians from the Spanish National Health Service. The email included the purpose of the questionnaire, the guidelines for its completion and an Excel file with the words selected. A second email was sent as a reminder in February 2018. The objective of this questionnaire was to obtain a score that represented how well these words are related to depression, as described by patients explaining their symptoms in clinical settings. The score of a word was obtained by adding the scores provided by each rater using a Likert scale (1. never, 2. rarely, 3. occasionally, 4. frequently and 5. very frequently). Since 20 raters participated in the scoring process, the maximum value of the score was 100 points. The list of 255 words and their scores are available at GitHub for research purposes: https://github.com/angelaleism/WordsDepression.

Finally, the third phase involved streaming of tweets that included at least one of the 375 words (the list of 255 words plus their different plural and gender forms), which was carried using the Twitter Application Programming Interface (API) [10]. The streaming was set up between June and September 2018 obtaining 8,832,256 tweets. In order to compare the usefulness of the words to detect tweets with signs of depression, two sets of 500 tweets were randomly selected after retweets removal. One of the sets was created by including five subsets of 100 tweets, each one composed by tweets that included one of the five most highly scored words, and the second 500-tweet set included low scored words. The 1,000 tweets were manually reviewed by two experts to determine whether they were potentially indicative of depression or not. The R programming language was used for the statistical analyses.

## 3. Results

The questionnaire was answered by 30% of psychiatrists (15/50) and all the family physicians (5/5; 100%). The respondents were 13 women and 7 men. All the respondents rated the complete list of 255 words. The mean and SD of the scores of all the participants are shown in figure 1.

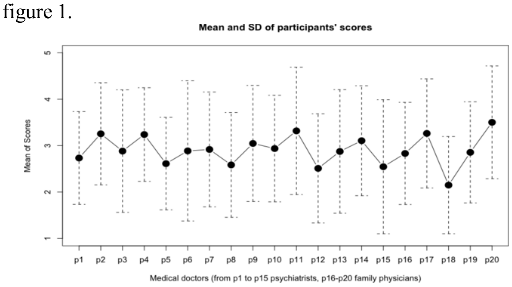



**Figure 1.** Distribution of the means and standard deviations (SD) of the 20 participants' scores

In order to assess the reliability of agreement among the raters, we calculated the intraclass correlation coefficients (ICCs) and the agreement among all the raters was 0.47, for the psychiatrists was 0.53 and for the family physicians was 0.37.

Based on the health professionals' scores, the 10 words in Spanish most frequently expressed by depressive patients were (the translation in English is in parenthesis): *deprimido/a* (depressed), *triste* (sad), *tristeza* (sadness), *desanimado/a* (downhearted), *depression* (depression), *depresivo/a* (depressive), *ansiedad* (anxiety), *cansado/a* (tired), *lloro* (crying), *insomnio* (insomnia). The 10 less frequent words were: *autocrítico* (self-criticism), *ingrato/a* (ungrateful), *mísero/a* (vile), *languidez* (languid), *mutilado/a* (disabled), *apetencia* (hunger), *sombrío/a* (gloomy), *achacoso/a* (sickly), *desdeñado/a* (disdained), *lasitud (*weariness). It is necessary to take into account that in the translation of these words into English there are some nuances that may be missing. There are some words that are more frequent in its feminine form on Twitter such as anxious, distrustful, distressed, insecure or shy and more frequent in masculine such as loser, solitary, incompetent or defeated.

Regarding the analysis of the tweets, table 2 shows the number of tweets that included the studied words, as well as the scores assigned to these words by the health professionals and the ranking of the words on the basis of the scores. In addition, the table shows the proportion of tweets for each word that were potentially indicative of depression when manually reviewed by an expert.

**Table 2.** Frequencies of words in the 8,832,256 tweets analysed, scores and proportion of depressive tweets

| Word | Number of tweets | Score (Rank) | % depressive tweets |
|---|---|---|---|
| Deprimido/a | 63,019 | 97 (1) | 12% |
| Triste | 56,776 | 97 (2) | 5% |
| Tristeza | 54,789 | 96 (3) | 10% |
| Desanimado/a | 24,079 | 94 (4) | 23% |
| Depresión | 51,408 | 92 (5) | 12% |
| Infelicidad | 12,116 | 63 (110) | 4% |
| Suicida | 42,615 | 57 (127) | 1% |
| Vencido/a | 43,020 | 52 (162) | 3% |
| Melancolía | 22,896 | 43 (194) | 9% |
| Desdichado/a | 4,663 | 41 (203) | 9% |

## 4. Discussion and conclusions

The diagnosis of depression is a complex process because of the heterogeneous nature of this disease, the lack of biological markers, the different symptoms among individuals and the diverse ways in which patients express those symptoms.

In relation with the scores assigned by the health professionals that participated in the survey, there was more agreement among the psychiatrists than among the family physicians. These results may be consistent with the fact that the psychiatrists deal with more patients with depressive disorders, and therefore, they are more familiar with the language used by patients with depression. As a result, the agreement between psychiatrists and family doctors is low. This finding may be due to the extraordinary diversity in the way people express themselves when describing their illness, the complexity of depression [9], how health professionals interpret these words for making a diagnosis and their clinical experience. For this reason, it is critical to perform a thorough analysis of the specific language used in each condition, before deciding the most suitable words to be used for filtering the tweets in each disease [2]. As our study shows, the words supposedly more linked to depression are very common words used in other contexts on Twitter, and consequently less specific for detecting depressive people (e.g. *triste*/sad). On the other hand, words less frequently mentioned on Twitter can have more weight to link them to depression or suicide tendency (e.g. *desanimado/* downhearted). The gender of words should be always considered when analyzing some languages that have grammatical gender forms such as Spanish.

Nevertheless, the analysis of words related to depression also require us to consider other aspects such as the linguistic features (i.e. the different use of personal pronouns, the number of negative words and the expressions associated to the basic emotions) and the behavioral patterns (i.e. the distribution of tweets over time and the number of characters or hashtags per tweet) [5]. However, the extensive list of words provided in this study can be used as a basis for developing new studies and strategies for the analysis of depression on Twitter in Spanish. Patients can be monitored, introducing new opportunities for studying depression and providing additional health services.

This study presents some limitations. On the one hand, the tweets can be considered an indirect and inaccurate way of detecting users suffering from depressive disorders and it is not possible to verify whether the diagnosis is genuine or not. On the other hand, although the list of words was carefully developed, there may be more expressions or words not included in the list.

## Acknowledgements

The Research Programme on Biomedical Informatics is a member of the Spanish National Bioinformatics Institute, PRB2-ISCIII, and it is supported by grant PT17/0009/0014. Special thanks to the psychiatrists of the Institute of Neuropsychiatry and Addiction (INAD) of Parc Salut Mar in Barcelona and the family community physicians who participated in the study.

## References

[1] C. Chung, J. Pennebaker. The Psychological Functions of Function Words. *Social Communication*, (2007), 343–359. Available at: http://doi.org/10.4324/9780203837702
[2] T. Nguyen, M. Larsen, D. Phung, S. Venkatesh, H. Christensen. Using linguistic and topic analysis to classify sub-groups of online depression communities. *Multimedia Tools and Applications* **76**, (2017), 10653–10676. Available at: Available from: http://doi.org/10.1007/s11042-015-3128-x
[3] W. Bucci, N. Freedman. The language of depression. *Bulletin of the Menninger Clinic* **45(4)**, (1981), 334–358.
[4] M. De Choudhury, M. Gamon, S. Counts, E. Horvitz. Predicting Depression via Social Media. In: *Proceedings of the Seventh International Conference on Weblogs and Social Media*, AAAI'13; July 8-11, 2013; Cambridge, MA p. 128-138.
[5] A. Leis, F. Ronzano, M.A. Mayer, L.I. Furlong, F. Sanz. Detecting signs of depression in tweets in Spanish: behavioral and linguistic analysis. *Journal of Medical Internet Research* **21(6)**, (2019), e14199.
[6] V.M. Prieto, S. Matos, M. Álvarez, F. Cacheda, J.L. Oliveira. Twitter: a good place to detect health conditions. *PLoS One* **9(1),** (2014), e86191.
[7] P.A. Cavazos-Rehg, M.J. Krauss, S. Sowles, S. Connolly, C. Rosas, M. Bharadwaj, et al. A content analysis of depression-related Tweets. *Computers in Human Behavior* **54**, (2016), 351-357
[8] Statista. Leading Countries Based on Number of Twitter Users as of January 2019 (in millions) (2019). Available at: https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/
[9] American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*. Washington, DC: American Psychiatric Publishing; 2013.
[10] Twitter Developer. Available at: https://developer.twitter.com/en.html

## 4.2. Detecting signs of depression in tweets in Spanish: behavioral and linguistic analysis

Social media platforms allow us to observe the activities, thoughts, and feelings of people's daily lives, including those of patients suffering from mental disorders. The objective of this study is to identify the linguistic features of tweets in Spanish and the behavioral patterns of Twitter users who generate them, which could suggest signs of depression. Depressive users are less active in posting tweets, doing it more frequently during the night. The first-person singular pronoun was by far the most used in the depressive users dataset. Emotions related to sadness, anger, and disgust were more common in the depressive users and depressive tweets datasets, when comparing with the control dataset. Negation words and negative polarity were more frequent in the depressive users and depressive tweets datasets than in the control dataset. Twitter users who are potentially suffering from depression modify the general characteristics of their language and the way they interact on social media. These users can be monitored and supported, thus introducing new opportunities for studying depression and providing additional health care services to people with this disorder.

Leis, A., Ronzano, F., Mayer, M. A., Furlong, L. I., & Sanz, F. (2019). Detecting signs of depression in tweets in Spanish: behavioral and linguistic analysis. *Journal of Medical Internet Research*, *21*(6), e14199. https://doi.org/10.2196/14199

Original Paper

# Detecting Signs of Depression in Tweets in Spanish: Behavioral and Linguistic Analysis

Angela Leis, PsyM; Francesco Ronzano, PhD; Miguel A Mayer, MD, PhD; Laura I Furlong, PhD; Ferran Sanz, Prof Dr

Research Programme on Biomedical Informatics, Hospital del Mar Medical Research Institute, Department of Experimental and Health Sciences, Universitat Pompeu Fabra, Barcelona, Spain

**Corresponding Author:**
Ferran Sanz, Prof Dr
Research Programme on Biomedical Informatics
Hospital del Mar Medical Research Institute
Department of Experimental and Health Sciences, Universitat Pompeu Fabra
Carrer Dr Aiguader 88
Barcelona, 08003
Spain
Phone: 34 933 160 540
Fax: 34 933 160 550
Email: ferran.sanz@upf.edu

## Abstract

**Background:** Mental disorders have become a major concern in public health, and they are one of the main causes of the overall disease burden worldwide. Social media platforms allow us to observe the activities, thoughts, and feelings of people's daily lives, including those of patients suffering from mental disorders. There are studies that have analyzed the influence of mental disorders, including depression, in the behavior of social media users, but they have been usually focused on messages written in English.

**Objective:** The study aimed to identify the linguistic features of tweets in Spanish and the behavioral patterns of Twitter users who generate them, which could suggest signs of depression.

**Methods:** This study was developed in 2 steps. In the first step, the selection of users and the compilation of tweets were performed. A total of 3 datasets of tweets were created, a depressive users dataset (made up of the timeline of 90 users who explicitly mentioned that they suffer from depression), a depressive tweets dataset (a manual selection of tweets from the previous users, which included expressions indicative of depression), and a control dataset (made up of the timeline of 450 randomly selected users). In the second step, the comparison and analysis of the 3 datasets of tweets were carried out.

**Results:** In comparison with the control dataset, the depressive users are less active in posting tweets, doing it more frequently between 23:00 and 6:00 ($P$<.001). The percentage of nouns used by the control dataset almost doubles that of the depressive users ($P$<.001). By contrast, the use of verbs is more common in the depressive users dataset ($P$<.001). The first-person singular pronoun was by far the most used in the depressive users dataset (80%), and the first- and the second-person plural pronouns were the least frequent (0.4% in both cases), this distribution being different from that of the control dataset ($P$<.001). Emotions related to sadness, anger, and disgust were more common in the depressive users and depressive tweets datasets, with significant differences when comparing these datasets with the control dataset ($P$<.001). As for negation words, they were detected in 34% and 46% of tweets in among depressive users and in depressive tweets, respectively, which are significantly different from the control dataset ($P$<.001). Negative polarity was more frequent in the depressive users (54%) and depressive tweets (65%) datasets than in the control dataset (43.5%; $P$<.001).

**Conclusions:** Twitter users who are potentially suffering from depression modify the general characteristics of their language and the way they interact on social media. On the basis of these changes, these users can be monitored and supported, thus introducing new opportunities for studying depression and providing additional health care services to people with this disorder.

***(J Med Internet Res 2019;21(6):e14199)*** doi: 10.2196/14199

**KEYWORDS**

depression; social media; mental health; text mining

## Introduction

### Background

Mental health is an essential component of our health. The World Health Organization (WHO) defines mental health as a "state of well-being in which people realize their potential, cope with the normal stresses of life, work productively, and contribute to their communities" [1]. Good mental health is about being cognitive, emotionally and socially healthy and it helps to determine the way we think and feel, in relation with others and how we make choices. Several factors, such as genetic, sociocultural, economic, political and environmental aspects, shape and influence our mental health. In the last few years, mental disorders have become a major concern in public health, and they are one of the main causes of the overall disease burden worldwide. They have devastating consequences for both patients and their families [2-7]. According to the WHO, depressive disorders are the most common among the mental illnesses [8]. Such disorders conditions are characterized by sadness, loss of interest and pleasure, feelings of guilt or low self-worth, disturbed sleep or appetite, feelings of tiredness, and poor concentration [8]. In 2018, at the global level, more than 300 million people were suffering from depression, and it is the main contributor to global disability. Depression has several consequences, both personal and social costs [9,10]. In some cases, depression can lead to suicide ideation and attempts [2,11]. The prevalence of this disorder changes depending on age, but it affects the whole population, from children and adolescents to elderly people. From 2005 to 2015, the number of people with depression increased by around 18% [12]. In this context, social media platforms allow to observe the activities, thoughts, and feelings of people's daily lives and thereby investigate their emotional well-being. This domain has become a new growing area of interest in public health and health care research [13-16]. People with depression often use social media to talk about their illness and treatment, share information and experiences, seek social support and advice, reduce social isolation, and manage their mental illness [15-21]. In addition, access to mobile devices facilitates the use of social media platforms, such as Twitter and Facebook, at any time and at any place. Social media, such as Twitter, is by nature social, and we can consequently find social patterns in Twitter feeds, thereby revealing key aspects of mental and affective disorders [22]. Social media has become an important source of health-related information, which allows us to detect and predict affective disorders and which can be used as an additional tool for mental health monitoring and infoveillance [23-26]. Furthermore, the application of different methodologies based on natural language processing and machine learning technologies has proved to be effective in supporting and automating the identification of early signs of mental illness by analyzing the content shared on the Web by individuals [13-15,27]. This human interaction with social media contributes to build the so-called digital phenotype, reshaping disease expression in terms of the lived experience of individuals and detecting early manifestations of several conditions [28]. Twitter is an internet microblogging social media service that allows users to post short messages about facts, feelings and opinions, and, as shown in previous studies, users' health conditions [15]. Twitter is one of the most important social media platforms in terms of number of users, with more than 330 million active users worldwide [29]. Since November 2017, the maximum number of characters of a tweet has been increased from 140 to 280. By analyzing huge amounts of text, researchers can link everyday language use with social behavior and personality [30,31]. Language, as a means of communication, constitutes an essential element for providing valuable insights about people's interests, feelings and concerns [32]. For this reason, the analysis of the messages posted on social media platforms may provide information about many personality traits, lifestyles, and psychological disorders [13,33,34]. The potential anonymity of social media encourages its users to be more willing to report health information, such as details of their mental disorders and treatments received. In addition, it is seen as a way to communicate and receive support from others with similar experiences, avoiding the isolation and fighting the social stigma of these conditions [12,15,17,19,32,35]. Nevertheless, users suffering from depression may also feel uncomfortable socializing and consuming information on social media platforms [36]. Several features of the messages, such as number and frequency of tweets, distribution throughout the day or during the night hours, and their seasonal character, can be used for the detection and monitoring of mental disorders, such as depression [20]. This knowledge can help health care professionals and health institutions and services in the decision-making processes to ensure better management of patients suffering from depression.

### Objectives

There are many studies that have used data mining and machine learning techniques on social media platforms to automatically identify people with mental health problems, such as depression, posttraumatic stress disorder, schizophrenia, or eating disorders, usually focusing the studies on messages written in English [20,37-39]. As far as we know, on social media, there are no studies about mental disorders that analyze messages written in Spanish. Taking into account that Spanish speaking countries, such as Spain and Mexico, are among the 10 most active Twitter users in the world, with more than 6 million and 7 million users, respectively [40], we focused our research on the expression of depression in Spanish language tweets. The aim of this study was to identify the linguistic features of tweets written in Spanish and the behavioral patterns of the corresponding Twitter users that could suggest signs of depression.

## Methods

### Study Steps

This study was designed and developed in 2 steps, with the aim of analyzing the linguistic patterns and behavioral features of Twitter users suffering from depression in comparison with the general population of Twitter users. The study was focused on tweets written in Spanish. In the first step, the selection of users and the compilation of tweets were performed. Given the design and purpose of the study, we decided to use the Twitter Application Programming Interface (API) [41]. Using this API, 3 datasets of tweets were created:

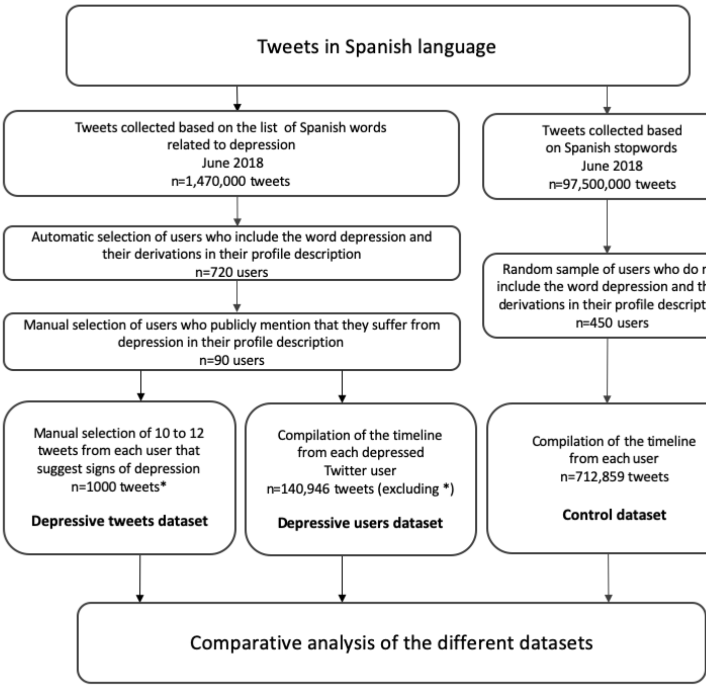
1. The *depressive users* dataset was made up of the timeline of 90 users who publicly mentioned on their Twitter profile that they suffer from depression.
2. The *control* dataset was made up of the timeline of 450 randomly selected Twitter users.
3. The *depressive tweets* dataset was constituted by a manual selection of tweets from the depressive users dataset, which specifically included expressions indicative of depression.

In the second step, comparison and analysis of the 3 datasets of tweets (control, depressive users, and depressive tweets datasets) were carried out to spot their distinguishing features. In the rest of this section, we will describe the methodology in detail. The flow diagram of the study is depicted in Figure 1.

### Data Collection and User Selection

The selection of the tweets and their users was based on the filtered real-time streaming support provided by the Twitter API. In the first step, we selected the users who showed potential signs of depression on Twitter on the basis of the 20 most frequent words in Spanish expressed by patients suffering from depression in clinical settings. These words were jointly identified and selected by a psychologist and a family physician with clinical experience and were based on the definition and general features of depression according to the Diagnostic and Statistical Manual of Mental Disorders [42]. The list of words used and their English translations are shown in Textbox 1.

**Figure 1.** Flow diagram of the study process.



**Textbox 1.** List of Spanish words related to depression and their English translations.

- agobiado/a (overwhelmed)
- agotado/a (exhausted)
- angustiado/a (distressed)
- ansiedad (anxiety)
- ansioso/a (anxious)
- cansado/a (tired)
- decaído (low)
- depresión (depression)
- depresivo/a (depressed as a condition)
- deprimido/a (depressed as state)
- desanimado/a (discouraged)
- desesperado/a (desperate)
- desmotivado/a (demotivated)
- insomnio (insomnia)
- llorar (cry)
- nervioso (nervous)
- preocupado/a (worried)
- solo/a (lonely)
- triste (sad)
- vacío/a (empty)

During June 2018, 1,470,000 tweets, including 1 or more occurrences of the words listed in Textbox 1, were collected. From this collection of tweets and to select the users who publicly stated in the textual description associated to their profile that they suffered from depression, all the profile descriptions, including 1 or more occurrences of the word "depr" and all the possible derivations related to the word depression in Spanish, such as "depre," "depresión," "depresivo," "depresiva," "deprimido," and "deprimida," were considered. From the 720 users who included 1 or more of these words in their description profile, 90 users who stated they suffered from depression or were receiving treatment for depression were selected for the analysis. This selection was performed by a psychologist, verifying that the statements were related to real expressions of depression, excluding quotes, jokes, or fake ones. For each of these depressed Twitter users, we collected all the most recent tweets from their timeline, up to a maximum of about 3200 tweets. Thus, a total of 189,669 tweets were collected, a figure that was reduced to 140,946 after discarding the retweets. These 140,946 tweets constituted the *depressive users dataset*. Examples of sentences appearing in the user profiles that were used for selecting the depressive users are:

- "Paciente psiquiátrico con depresión crónica" (*Psychiatric patient with chronic depression;* example of a profile sentence that indicates depression).
- "Colecciono errores traducidos a tweets depresivos y a uno que otro impulso de amor" (*I gather errors translated into depressing tweets and into one or another love impulse*; example of a profile sentence that does not indicate depression).

Once the users with profile sentences indicating depression had been retrieved, their Twitter timelines were collected. Only those users having in their timeline at least 10 tweets that suggested signs of depression were retained for further analyses. For each user, the selection of these tweets was performed by manually inspecting the tweets of the user's complete timeline in reverse temporal order, starting from the most recent one to the oldest tweet of the timeline retrieved by means of the Twitter API . Finally, a total number of 1000 tweets issued by the 90 depressive users, suggesting signs of depression, were detected and used for the analysis. This set of tweets provided us with the *depressive tweets dataset*, which was used to analyze linguistic features of tweets showing signs of depression. It has to be mentioned that these 1000 tweets were not to be included in the depressive users dataset (see Figure 1). At the same time, more than 97,500,000 tweets were also collected in June 2018: such tweets were gathered by listening to the public Twitter stream during this time span by only considering tweets with Spanish textual contents (as detected by Twitter language identification support).

Given that Twitter requires more restrictive filters than just the language of the tweets, we used a list of the most frequently used Spanish words (stopwords) to retrieve all tweets that included 1 or more of these words. The vast majority of Spanish tweets should match this criterion. A sample of 450 users who did not mention in their profile the word depression and its derivations were selected randomly from the 97,500,000 tweets. The complete timelines of these users were compiled (1,141,021 tweets), which were reduced to 712,589 once retweets were removed. These 712,589 tweets constituted the *control dataset*. To identify the language of a tweet, we relied on the language automatically identified by Twitter for each tweet, selecting tweets in Spanish. It has to be noted that these data can contain some tweets from unidentified depressive users.

### Data Analysis

A comparison of the 3 datasets was performed to determine the existence of differential linguistic and behavioral features. The different features that were analyzed are shown in Table 1.

The textual content of each tweet was analyzed by means of the following sequence of steps:

- Tokenization performed by means of a custom Twitter tokenizer included in the Natural Language Toolkit [43].
- Part-of-Speech (POS) tagging performed by means of the Freeling Natural Language Processing tool in order to analyse the usage patterns of grammatical categories (eg, adjectives, nouns, or pronouns) in the text of tweets [44].
- Identification of negations performed by relying on a custom list of Spanish negation expressions, such as *nada* (nothing), *nadie* (nobody), *no* (no), *nunca* (never), and alike.
- Identification of occurrences of positive and negative words inside the text of each tweet by means of 2 Spanish polarity lexicons: the Spanish Sentiment Lexicon and the Spanish SentiCon Lexicon [45,46]. We exploited 2 lexicons to consider and compare 2 approaches of modeling polarity in Spanish texts, thus reducing any language modeling bias that the use of a single language resource could introduce.
- Identification of words and expressions associated to the basic emotions [47] by using the Spanish Emotion Lexicon [48]. Such emotions are *alegría* (happiness), *enojo* (anger), *miedo* (fear), *repulsión* (disgust), *sorpresa* (surprise), and *tristeza* (sadness).

All the tools and aforementioned resources are publicly available. The statistical analyses were carried out with the R version 3.4.3 (R Development Core Team) and SPSS Statistics version 23.0 (IBM), applying the relevant test for each type of comparison to be carried out.

**Table 1.** Characteristics of the tweets analyzed.

| Feature | Analyses performed |
|---|---|
| Distributions over time | • Tweets throughout the day (per hour)<br>• Tweets throughout the week |
| Part-of-Speech | • Number of words by grammatical categories (part-of-speech tags)<br>• Number of personal pronouns |
| Counts | • Number of characters<br>• 200 most frequent words (word cloud)<br>• Number of hashtags, links, mentions, and emojis |
| Emotion analysis | • Emotion types and frequencies |
| Negations | • Negation words types and frequencies |
| Polarity analyses | • Polarity of tweets on the basis of Spanish Sentiment Lexicon and Spanish SentiCon Polarity |

### Ethical Approval

The protocol used in this study was approved by the Ethics Committee of Parc Salut Mar (approval number 2017/7234/1).

## Results

### Distribution Over Time

Regarding the distribution of tweets over time, the number of tweets per hour and throughout the week of control and depressive users datasets were compared. The tweet hours were adjusted by the user's time zone. As shown in Figure 2, the depressive users are less active in generating tweets than the control ones, reaching both groups the same activity level between 23:00 and 6:00. The comparison of the temporal distributions of tweets between both datasets was carried out by means of a repeated measures analysis of variance (Greenhouse-Geisser F=6.605; $P$<.001). As shown in Figure 3, the activity throughout the week of the depressive users dataset presented more regular activity than the control dataset, whose users' activity showed a sharp drop during the weekend. The differences between these datasets were statistically significant (Greenhouse-Geisser F=4.153; $P$=.008).

**Figure 2.** Number of tweets and retweets per hour of the control and depressive users datasets (mean±standard error of mean). SEM: standard error of mean.

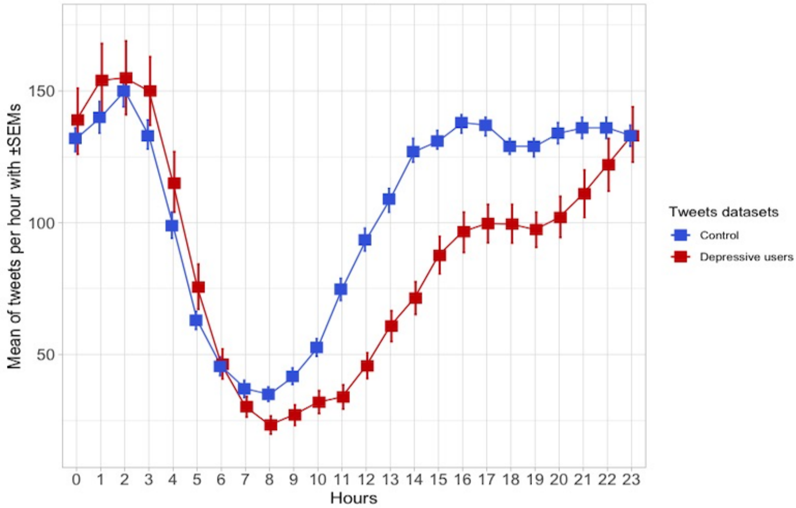
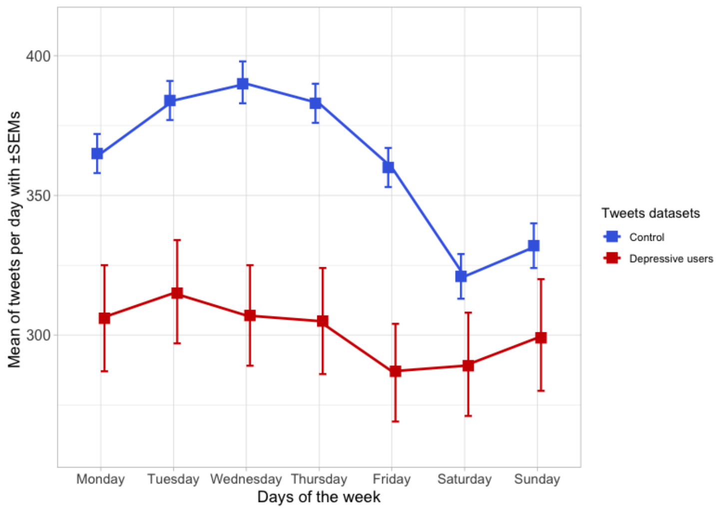



**Figure 3.** Number of tweets and retweets throughout the week of the control and depressive users datasets (mean±standard error of mean). SEM: standard error of mean.



### Part-of-Speech

As for the analysis of POS corresponding to the number of words by grammatical categories in each tweet, we compared the 3 datasets of tweets: the control, depressive users, and depressive tweets datasets. As previously stated, the tweets of the depressive tweets dataset were removed from the depressive users dataset. The frequencies of words in each group are shown in Table 2. The number of nouns used in the control group almost doubles that of the depressive users dataset. By contrast, verbs are more frequently used in the depressive users dataset than in the control dataset. There were statistically significant differences between the control and the depressive users datasets ($\chi^2_7$=1,242,600; $P$<.001), between the control and the depressive tweets datasets ($\chi^2_7$=2,105.7; $P$<.001), and between the depressive users and the depressive tweets datasets ($\chi^2_7$=15,888; $P$<.001).

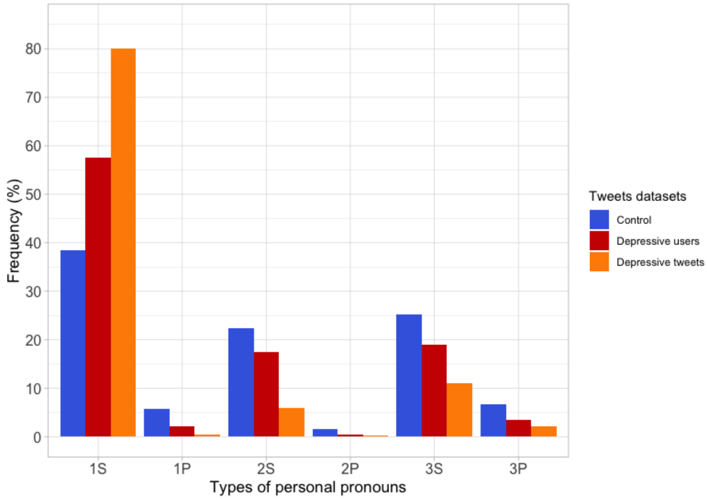
In relation to the different types of pronouns in the control dataset, we detected 396,181 personal pronouns (51.38%; 396,181/770,955), the first-person singular (38.37%; 152,013/396,181) being the most used. A similar profile was observed in the depressive users dataset, where 124,614 pronouns were found (55.16%; 124,614/225,913), the first-person singular remaining the most used (57.59%; 71,768/124,614). In the depressive tweets dataset, 865 personal pronouns (53.16%; 865/1,627) were identified, and the first-person singular pronoun was by far the most used (80.00%; 692/865). The frequencies of personal pronouns in the different datasets are shown in Figure 4. There were statistically significant differences between the control and the depressive users datasets ($\chi^2_5$=15,912; $P$<.001), between the control and the depressive tweets datasets ($\chi^2_5$=638.7; $P$<.001), and between the depressive users and the depressive tweets datasets ($\chi^2_5$=183.9; $P$<.001).

In relation to the number of characters per tweet, the mean of characters per tweet in the control and depressive users datasets was 83.48 (SD 40.57) and 65.76 (SD 36.99) characters, respectively, with statistically significant differences between them ($t_{213770}$=161.6; $P$<.001). On the other hand, the mean in the depressive tweets dataset was 67.51 (SD 38.28), which was not statistically significant and different in comparison with the depressive dataset ($t_{1012.3}$=1.45; $P$=.15). The 200 most frequent words that appeared in the control and depressive users datasets are depicted in the 2 word clouds shown in Multimedia Appendix 1. The 10 most frequent words that appeared in the control dataset were the following: *hoy* (today), *día* (day), *ver* (to see), *quiero* (I want), *gracias* (thank you), *mejor* (better), *siempre* (always), *vida* (life), *ahora* (now), and YouTube. In the depressive users dataset, the 10 most frequent words were the following: *quiero* (I want), *vida* (life), *siempre* (always), *siento* (I feel), *nadie* (nobody), *mierda* (shit), *never* (nunca), and *día* (day). It should be noted that in the depressive tweets dataset, although there are several words in common with the depressive users dataset, we can find additional words that are not present in the other datasets, such as *vacío/a* (empty), *matar* (to kill), *desaparecer* (to disappear), *suicidar* (commit suicide), *muerta* (dead), *desastre* (disaster), *inútil* (useless,), *deprimida* (depressed as state in women), *depresiva* (depressed as a condition in women), and *insomnio* (insomnia). The word cloud of the depressive tweets dataset is shown in Multimedia Appendix 2. In relation to the use of links, hashtags, and mentions in tweets, the frequency of them in the control and depressive users datasets were 35.32% (251,728/712,584), 13.13% (93,575/712,588), and 44.00% (313,574/712,577) and 18.07% (25,475/140,946), 1.44% (2030/140,946), and 9.27% (13,060/140,942), respectively. The number of tweets, including emojis, were 13.61% (97,038/712,589) in the control dataset and 5.72% (8069/140,947) in the depressive users dataset.

**Table 2.** Part-of-Speech (POS) frequencies in tweets of control, depressive users, and depressive tweets datasets.

| Type of POS | POS in the control dataset, n (%) | POS in the depressive users dataset, n (%) | POS in the depressive tweets dataset, n (%) |
|---|---|---|---|
| Noun | 2,298,544 (28.48) | 270,104 (17.77) | 1776 (15.07) |
| Verb | 1,660,700 (20.58) | 400,755 (26.36) | 3391 (28.77) |
| Pronouns | 770,955 (9.55) | 225,913 (14.86) | 1627 (13.80) |
| Adjectives | 593,327 (7.35) | 83,089 (5.47) | 588 (4.99) |
| Determiner | 1,068,130 (13.23) | 177,795 (11.70) | 1342 (11.39) |
| Adverbs | 496,988 (6.16) | 140,963 (9.27) | 1351 (11.46) |
| Adpositions | 854,573 (10.59) | 123,867 (8.15) | 1052 (8.93) |
| Conjunctions | 327,852 (4.06) | 97,541 (6.42) | 659 (5.59) |

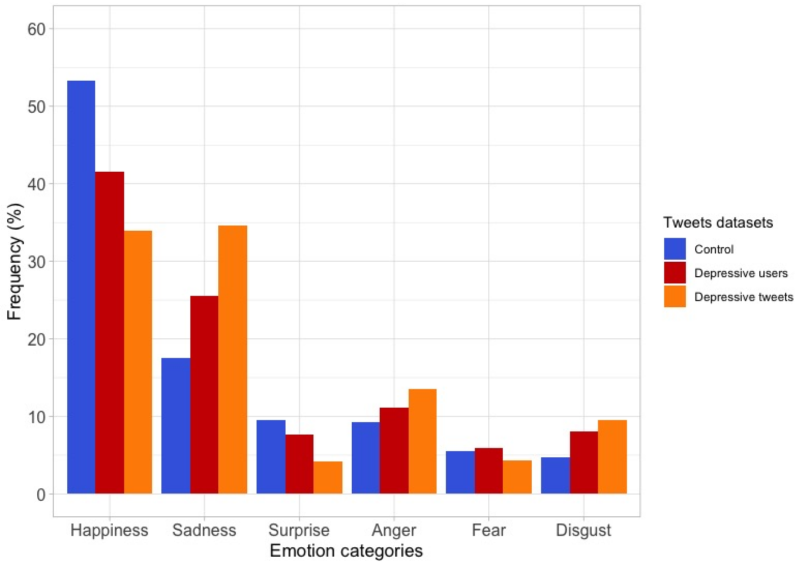
**Figure 4.** Frequency of the different types of personal pronouns in the control, depressive users, and depressive tweets datasets. 1S: first-person singular; 1P: first-person plural; 2S: second-person singular; 2P: second-person plural; 3S: third-person singular; 3P: third-person plural.




## Emotion Analysis

Regarding the distribution of emotions, in the control dataset and in the depressive users dataset, the most frequent emotion was happiness (53.30%; 203,029/380,874 and 41.60%; 40,535/97,425) followed by sadness, which was more frequent in the depressive users dataset (17.59%; 67,033/380,874 and 25.49%; 24,834/97,425). In the depressive tweets dataset, the most frequent emotion was sadness (34.00%; 303/891). There were statistically significant differences between the control and the depressive users datasets ($\chi^2_5$=6838.2; $P$<.001), between the control and the depressive tweets datasets ($\chi^2_5$=296.8; $P$<.001), and between the depressive users and the depressive tweets datasets ($\chi^2_5$=65.6; $P$<.001). The frequencies of the different emotions are shown in Figure 5.

**Figure 5.** Frequency distributions of emotion categories in the tweets of the 3 datasets.



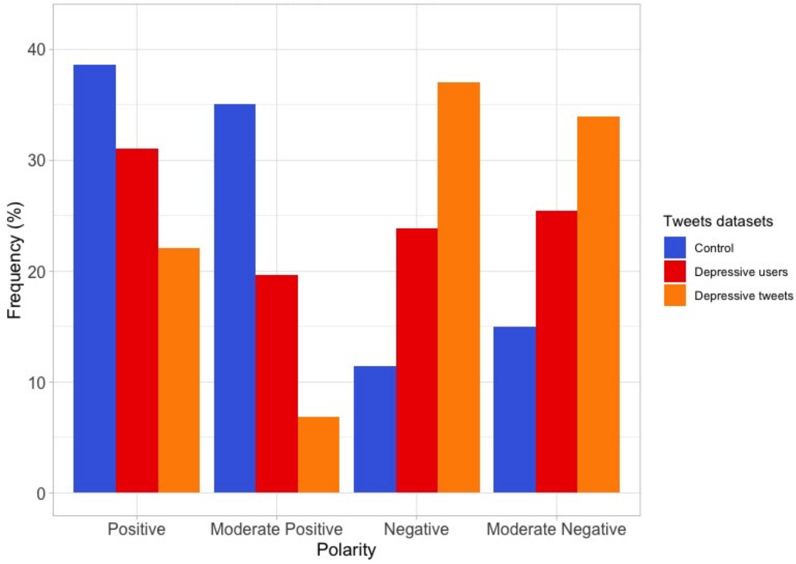

### Negation Words

Regarding the use of negation words, they were detected in 21.74% (154,953/712,588) of the tweets in the control dataset, in 34.15% (48,137/140,946) of the depressive users dataset, and in 45.50% (455/1000) of the depressive tweets dataset. The mean of negation words was 0.28 (SD 0.59) in the control dataset, it was 0.49 (SD 0.82) in the depressive users dataset, and it was 0.67 (SD 0.91) in the depressive tweets dataset. There were statistically significant differences between the control and the depressive users datasets (Mann-Whitney $U$=4.3657e+10; $P$<.001), between the control and the depressive tweets datasets (Mann-Whitney $U$=266,990,000; $P$<.001), and between the depressive users and the depressive tweets datasets (Mann-Whitney $U$=62,002,000; $P$<.001).

### Polarity Analysis

In relation to the polarity of tweets, 2 analyses were performed using 2 Spanish sentiment lexicons: the Senti Lexicon (including positive and negative categories) and the SentiCo Polarity (including positive, moderate positive, moderate negative, and negative categories). According to the Senti Lexicon, the analysis of tweets showed that the control dataset shows polarity in 33.47% (245,367/733,029) of the tweets, being positive in 56,54% (138,726/245,367) of them. In contrast, the depressive users dataset shows polarity in 41,31% (61,132/147,996) of the tweets, being positive in 46.14% (28,205/61,132) of them. Finally, the depressive tweets dataset shows polarity in 58.90% (589/1000) of the tweets, with positive polarity in 34.97% (206/589) of them. There were statistically significant differences between the control and the depressive users datasets ($\chi^2_1$=2134; $P$<.001), between the control and the depressive tweets datasets ($\chi^2_1$=110.3; $P$<.001), and between the depressive users and the depressive tweets datasets ($\chi^2_1$=28.8; $P$<.001). When using the SentiCo Polarity tool, the control dataset presented 20.97% (152,228/725,717) of tweets with polarity, 29.32% (42,820/146,033) in the depressive users and 33.34% (348/1,044) in the depressive tweets dataset. The distributions of polarities are shown in Figure 6. There were statistically significant differences between the control and the depressive users datasets ($\chi^2_3$=8820.8, $P$<.001), between the control and the depressive tweets datasets ($\chi^2_3$=308.8; $P$<.001), and between the depressive users and the depressive tweets datasets ($\chi^2_3$=52.4; $P$<.001).

**Figure 6.** Polarities of the tweets according to the SentiCo Polarity tool in the 3 datasets.




## Discussion

### Principal Findings

The diagnosis of depression is complex because of the heterogeneous nature of this disease and the diverse manifestation of the symptoms among individuals, which result in a great number of depressive disorder cases that are undetected and untreated, making the prevention, diagnosis, and treatment of the depressive disorders a complicated task [15,49,50]. For these reasons and taking into account that people diagnosed with depression are increasing worldwide, new strategies for detecting and monitoring this disease would be very useful. In this study, we analyzed the behavioral and linguistic patterns of tweets in Spanish that suggest signs of depression. The results contribute to the growing body of scientific literature that analyzes the messages posted on social media using languages other than English. We have introduced a new approach that comprises analyzing the timelines of self-qualified depressed users, as well as their tweets related to depression, which are manually selected. Our results show that the tweets of depressive users have different features in comparison with those of a control dataset, even when their tweets that are not related to depression are considered (depressive users dataset). In addition, the differences with the control dataset become more evident when we consider the manual selection of tweets related to depression (depressive tweets dataset).

### Different Distributions of Tweets Over Time

As for the distribution of tweets over time, the users of the depressive dataset, although being less active in using Twitter, used to be more active during the night than the users of the control dataset. This can be explained as a result of insomnia, one of the most frequent symptoms of depression. This finding is consistent with previous studies carried out with English speakers, which demonstrated that individuals with depression are more active during the night [20]. Moreover, the daily mood changes, such as the morning and evening worsening that are typical in several forms of depression, could explain the lower activity of the depressive users [51]. In relation to the distribution of tweets throughout the week, the users of the depressive dataset showed a more regular activity throughout the week, tending to be more active on Saturdays, Sundays, and Mondays than those of the control dataset, whose activity showed a drop during the weekend. This trend may be related to the lowered social activity of the people suffering from depressive disorders, having a reduced participation in social leisure activities during the weekend and spending more time at home, sharing their feelings and thoughts on social media platforms [16].

### Different Style of Writing

The analysis of POS and the number of words by grammatical categories show that, generally, the users of the depressive dataset used more verbs, adverbs, and pronouns but less nouns than the control dataset. The same features are also present in the depressive tweets dataset. These findings suggest that the language of people suffering from depression is characterized by a different style of writing that some authors describe as poorly structured, indicating less interest in what surrounds them, people, objects, or things [52]. They focus on talking about actions, and this is correlated with sensitive disclosure. Consistent with many previous studies [20,30,35,53-55], the use of first-person singular pronoun is more frequent among the users of the depressive dataset, with respect to those of the control dataset, and this difference increases in the depressive tweets dataset. The increased use of this pronoun demonstrates the attention to self-focus that is associated with the negative emotional states of depression and the reduced attentional resources, highlighting the psychological distancing to connect with others [56]. This social isolation may also explain that the first- and second-person plural pronouns are the least used. Language can be used as a measure of different individual features, on the basis of the fact that people's word choice is stable over time and consistent across topics or context. For this reason, the language style appears to be a useful predictor of some mental health conditions, such as depression [20,35]. In addition, the number of characters written in the depressive users and depressive tweets datasets was smaller than the number of characters written in the control dataset, and this might be related to reduced interest and poorer language. According to the most frequent words that appeared in the depressive users and depressive tweets datasets, there are specific words that are linked to clinical symptoms and the way that depressive patients word their mood, such as words that may be related to suicide ideation. Consequently, they can be used as a signal to detect potentially depressed users on Twitter [36]. Similarly, we observed the frequent use of adjectives in feminine form in the depressive tweets dataset, which would suggest that a high proportion of the depressive users are women, a fact that is in agreement with clinical and epidemiological evidences [8,11,12,42].

### Predominant Emotions

Emotions are one of the key aspects that characterize many mental health conditions and, particularly, when people are suffering from depression. An analysis of the 6 emotions that are commonly considered (happiness, sadness, surprise, anger, fear, and disgust) [47] was performed to determine the existence of differences among the datasets. Happiness is the most frequent emotion in the control and depressive users datasets, although an important reduction was observed in the depressive tweets dataset. The surprise emotion is less frequent in depressive users and, specially, in the depressed tweets datasets than the control dataset, and this fact can be related to the depressive mood, in which there is a decrease in interest in almost everything. Fear does not seem to be a differential emotion in the groups of tweets analyzed in this study.

Regarding negative emotions, we observed an increase in the frequency of words related to the sadness emotion in the depressive tweets dataset, doubling that of the control dataset. This feature had also been observed in other studies [14,35,57]. Moreover, anger is more frequent in the depressive user and depressive tweets datasets than in the control dataset. Although Twitter is used many times for expressing anger about personal or political aspects, this emotion is particularly frequent in patients suffering from depression, who tend to feel irritable, wronged, or angry at the world [14,16,35,58]. At the same time, disgust, an emotion that is known to be associated with the depressive disorders [59], was found to be more frequent in the depressive users and depressive tweets datasets.

### Negative Focused Emotion Language

In our analysis, the presence of negation words is more frequent in the depressed users (34.15%; 48,137/140,946) and depressive tweets (45.50%; 455/1000) datasets than in the control dataset (21.74%; 154,953/712,588), indicating that there is an increased use of negatively focused emotion language, which is typical in depressive patients and feelings [31,54,55,60].

### Negative Polarity

The classification of tweets, on the basis of the emotional positivity or negativity of their words, is another analysis that has been carried out. In this study, we used 2 types of polarity lexicons, the Senti Lexicon (SentiLex) and the Sentico Polarity (SentiCo). In both cases, the negative polarity was higher in the depressive users and depressive tweets datasets, even tripling the negativeness of the control dataset when using the Sentico Polarity lexicon. These findings are consistent with other studies, indicating that people suffering from depression tend to focus more on negative aspects of their life [20,35], and thus their tweets contain much more negative emotional words compared with the control dataset [14]. In addition, the self-focus state that characterizes depression is associated with negative emotions [32,56,57].

### Limitations and Future Directions

This study presents some limitations that have to be pointed out. On the one hand, the tweets of the depressive datasets come only from Twitter users who speak publicly about feelings and emotions that can be related with depression. This is an indirect and inaccurate way of detecting users suffering from depressive disorders. Without clinically assessing these people, there is no way to verify if the diagnosis is genuine or if they suffer from another mental disorder. On the other hand, it is possible that Twitter users self-disclose their mental health using words or expressions not included in the list of keywords used in this study for streaming tweets about depression [22,61-63]. In this respect, it is possible that a wider list could have yielded a greater coverage [21,36]. Privacy policies of social media restrict the access to users who did not grant access to their profile, and this may have generated biases in the composition of the depressive users and the depressive tweets datasets. In addition, tweets may incorporate biases because of the self-management and anonymity of the Web-based identities [61]. Moreover, Twitter users may be not be representative of the general population, and some studies have shown that they are often urban people with high levels of education [64-66]. More information about the socioeconomic and demographic details of Twitter users is needed [67]. The control dataset was a randomly selected sample of Twitter users, and it is consequently representative of the users of this social media. However, there is a possibility that users in this group may also have depression or other mental illness even though they did not mention this in their profile description. There is also the possibility that the users included in the control group are fake accounts. Only original tweets were analyzed, and perhaps retweets, which are not included in our linguistic study, reflect users' emotions that can be related to depression status [68]. Finally, depression is a very complex mental disorder, and our study only provides a general observation of this disorder. Additional research might be carried out to examine specific depression types and determine if there are social media features that can contribute to classifying users or tweets to the different diagnosis of depression [69]. Similarly, in future works, we plan to study the linguistic features and the behavioral patterns of depression in different linguistics contexts. The possible relationship between depression and seasonality could be of interest for future studies in the context of monitoring Twitter activity [70].

### Conclusions

The prevalence of common mental disorders worldwide, such as depression, requires the ability of health care systems to provide adequate diagnosis, monitoring, and treatment. The wide popularity of social media platforms introduces new opportunities for the screening of depression. The introduction of new methods of analysis for the automatic detection of signals of depression on social media platforms, such as Twitter or Facebook, has the potential of being used as a complementary tool for the assessment of these patients, assisting health care professionals in the detection and monitoring of mental health disorders. Although the analysis of tweets as a way to determine the existence of depression cannot be used as a replacement for diagnosis, it has the potential as a screening tool for depressive disorders, with a lower cost than other traditional procedures. In addition, it can be helpful to health professionals for managing and monitoring patients more efficiently. Similarly, it can be useful for particular patients, as they feel more comfortable disclosing their symptoms on Twitter than in clinical settings. In this study, we have shown that several behavioral and linguistic features of the tweets in Spanish can be used as a complementary tool to detect signals of depression of their authors, corroborating and extending the findings obtained by studies carried out on English tweets. As we described in this study, signs of depression of Twitter users are not exclusively spotted by identifying and analyzing tweets that explicitly mention expressions related to depression. Moreover, Twitter users who are potentially suffering from depression globally modify the core traits of their language, independently from the fact that the tweets are related or not related to the expression of depression. On the basis of these changes, these users can be monitored and supported. The results of this paper, jointly with other studies on the matter, support the potential of social media as an important instrument for extending and enhancing mental health services available to people with mental disorders. By means of interdisciplinary collaborations, it is possible to develop digital apps and services providing personalized alerts and psychosocial support in the mental health domain.

### Acknowledgments

We received support from the Agency for Management of University and Research Grants in Catalonia (Spain) for the incorporation of new research personnel (FI2016), the European Union H2020 Research and Innovation Programme 2014-2020 under grant agreement number 634143 (MedBioinformatics: Creating medically driven integrative bioinformatics applications focused on oncology and central nervous system disorders and their comorbidities). The Research Programme on Biomedical Informatics is a member of the Spanish National Bioinformatics Institute, PRB2-ISCIII, and it is supported by grant PT17/0009/0014.

### Conflicts of Interest

None declared.

### Multimedia Appendix 1

Word clouds showing the 200 most frequent words in the control (left) and depressive users (right) datasets.

[PNG File, 396KB-Multimedia Appendix 1]

### Multimedia Appendix 2

Word cloud showing the 200 most frequent words in the depressive tweets dataset.

[PNG File, 166KB-Multimedia Appendix 2]

### References

1. World Health Organization. 2013. Mental Health Action Plan 2013-2020 URL:http://apps.who.int/iris/bitstream/10665/89966/1/9789241506021_eng.pdf?ua=1 [accessed 2018-04-25] [WebCite Cache ID 778TsWziL]
2. Marcus M, Yasamy MT, van Ommeren M, Chisholm D, Saxena S, WHO Department of Mental Health and Substance Abuse. World Health Organization. 2012. Depression: A Global Public Health Concern URL:http://www.who.int/mental_health/management/depression/who_paper_depression_wfmh_2012.pdf [accessed 2018-04-25] [WebCite Cache ID 778NmzRq1]
3. Global Burden of Disease Study 2013 Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: a systematic analysis for the

global burden of disease study 2013. Lancet 2015 Aug 22;386(9995):743-800 [FREE Full text] [doi: 10.1016/S0140-6736(15)60692-4] [Medline: 26063472]

4. Trautmann S, Rehm J, Wittchen HU. The economic costs of mental disorders: do our societies react appropriately to the burden of mental disorders? EMBO Rep 2016;17(9):1245-1249 [FREE Full text] [doi: 10.15252/embr.201642951] [Medline: 27491723]
5. Patel V, Chisholm D, Parikh R, Charlson FJ, Degenhardt L, Dua T, et al. Global priorities for addressing the burden of mental, neurological,substance use disorders. In: Patel V, Chisholm D, Dua T, Laxminarayan R, Medina-Mora ME, Vos T, editors. Disease Control Priorities: Mental, Neurological, and Substance Use Disorders, Third Edition (Volume 4). Washington, DC: The World Bank; 2016:1-27.
6. Whiteford HA, Degenhardt L, Rehm J, Baxter AJ, Ferrari AJ, Erskine HE, et al. Global burden of disease attributable to mental and substance use disorders: findings from the global burden of disease study 2010. Lancet 2013 Nov 9;382(9904):1575-1586. [doi: 10.1016/S0140-6736(13)61611-6] [Medline: 23993280]
7. Wongkoblap A, Vadillo MA, Curcin V. Researching mental health disorders in the era of social media: systematic review. J Med Internet Res 2017 Dec 29;19(6):e228 [FREE Full text] [doi: 10.2196/jmir.7215] [Medline: 28663166]
8. World Health Organization. 2018. Depression: Key Facts URL:https://www.who.int/news-room/fact-sheets/detail/depression [accessed 2018-04-23] [WebCite Cache ID 778TC0BMP]
9. Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. PLoS Med 2006 Nov;3(11):e442 [FREE Full text] [doi: 10.1371/journal.pmed.0030442] [Medline: 17132052]
10. Vigo D, Thornicroft G, Atun R. Estimating the true global burden of mental illness. Lancet Psychiatry 2016 Feb;3(2):171-178. [doi: 10.1016/S2215-0366(15)00505-2] [Medline: 26851330]
11. World Health Organization. 2017. Depression and Other Common Mental Disorders: Global Health Estimates URL:http://apps.who.int/iris/bitstream/10665/254610/1/WHO-MSD-MER-2017.2-eng.pdf?ua=1 [accessed 2018-04-25] [WebCite Cache ID 77CsSorTr]
12. Ferrari AJ, Charlson FJ, Norman RE, Patten SB, Freedman G, Murray CJ, et al. Burden of depressive disorders by country, sex, age, and year: findings from the global burden of disease study 2010. PLoS Med 2013 Nov;10(11):e1001547 [FREE Full text] [doi: 10.1371/journal.pmed.1001547] [Medline: 24223526]
13. Paul MJ, Dredze M. You Are What You Tweet: Analyzing Twitter for Public Health. In: Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media. 2011 Presented at: AAAI'11; July 17-21, 2011; Barcelona, Spain p. 265-272.
14. Park M, Cha C, Cha M. Depressive Moods of Users Portrayed in Twitter. In: Proceedings of the ACM SIGKDD Workshop on Health Informatics. 2012 Presented at: HI-KDD'12; August 12-16, 2012; Beijing, China p. 1-8.
15. Nguyen T, O'Dea B, Larsen M, Phung D, Venkatesh S, Christensen H. Using linguistic and topic analysis to classify sub-groups of online depression communities. Multimed Tools Appl 2015 Dec 21;76(8):10653-10676. [doi: 10.1007/s11042-015-3128-x]
16. Cavazos-Rehg PA, Krauss MJ, Sowles S, Connolly S, Rosas C, Bharadwaj M, et al. A content analysis of depression-related Tweets. Comput Human Behav 2016;54:351-357 [FREE Full text] [doi: 10.1016/j.chb.2015.08.023] [Medline: 26392678]
17. Coppersmith G, Dredze M, Harman C, Hollingshead K. From ADHD to SAD: Analyzing the Language of Mental Health on Twitter Through Self-Reported Diagnoses. In: Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality. 2015 Presented at: CLPSYCH'15; June 5, 2015; Denver, Colorado p. 1-10 URL:https://www.aclweb.org/anthology/W15-1201 [doi: 10.3115/v1/W15-1201]
18. Naslund JA, Aschbrenner KA, McHugo GJ, Unützer J, Marsch LA, Bartels SJ. Exploring opportunities to support mental health care using social media: a survey of social media users with mental illness. Early Interv Psychiatry 2019 Jun;13(3):405-413. [doi: 10.1111/eip.12496] [Medline: 29052947]
19. Berry N, Lobban F, Belousov M, Emsley R, Nenadic G, Bucci S. #WhyWeTweetMH: understanding why people use Twitter to discuss mental health problems. J Med Internet Res 2017 Dec 5;19(4):e107 [FREE Full text] [doi: 10.2196/jmir.6173] [Medline: 28381392]
20. De Choudhury M, Gamon M, Counts S, Horvitz E. Predicting Depression via Social Media. In: Proceedings of the Seventh International Conference on Weblogs and Social Media. 2013 Presented at: AAAI'13; July 8-11, 2013; Cambridge, MA p. 128-138.
21. Wilson ML, Ali S, Valstar MF. Finding Information About Mental Health in Microblogging Platforms: A Case Study of Depression. In: Proceedings of the 5th Information Interaction in Context Symposium. 2014 Presented at: IIiX'14; August 26-30, 2014; Regensburg, Germany p. 8-17. [doi: 10.1145/2637002.2637006]
22. Coppersmith G, Dredze M, Harman C. Quantifying Mental Health Signals in Twitter. In: Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality. 2014 Presented at: CLPSYCH'14; June 5, 2014; Baltimore, Maryland p. 51-60. [doi: 10.3115/v1/W14-3207]
23. Radzikowski J, Stefanidis A, Jacobsen KH, Croitoru A, Crooks A, Delamater PL. The measles vaccination narrative in Twitter: a quantitative analysis. JMIR Public Health Surveill 2016;2(1):e1 [FREE Full text] [doi: 10.2196/publichealth.5059] [Medline: 27227144]

24. Stefanidis A, Vraga E, Lamprianidis G, Radzikowski J, Delamater PL, Jacobsen KH, et al. Zika in Twitter: temporal variations of locations, actors, and concepts. JMIR Public Health Surveill 2017 Apr 20;3(2):e22 [FREE Full text] [doi: 10.2196/publichealth.6925] [Medline: 28428164]
25. Finch KC, Snook KR, Duke CH, Fu K, Tse ZT, Adhikari A, et al. Public health implications of social media use during natural disasters, environmental disasters, and other environmental concerns. Nat Hazards 2016 Apr 19;83(1):729-760. [doi: 10.1007/s11069-016-2327-8]
26. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the internet. J Med Internet Res 2009 Mar 27;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
27. Conway M, O'Connor D. Social media, big data, and mental health: current advances and ethical implications. Curr Opin Psychol 2016 Jun;9:77-82 [FREE Full text] [doi: 10.1016/j.copsyc.2016.01.004] [Medline: 27042689]
28. Jain SH, Powers BW, Hawkins JB, Brownstein JS. The digital phenotype. Nat Biotechnol 2015 May;33(5):462-463. [doi: 10.1038/nbt.3223] [Medline: 25965751]
29. Statista. 2018. Number of Monthly Active Twitter Users Worldwide From 1st Quarter 2010 to 2nd Quarter 2018 (in Millions) URL:https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/ [accessed 2018-10-20] [WebCite Cache ID 778OhlXHI]
30. Tausczik YR, Pennebaker JW. The psychological meaning of words: LIWC and computerized text analysis methods. J Lang Soc Psychol 2009 Dec 8;29(1):24-54 [FREE Full text] [doi: 10.1177/0261927X09351676]
31. Ramirez-Esparza N, Chung CK, Kacewicz E, Pennebaker JW. The Psychology of Word Use in Depression Forums in Englishin Spanish: Texting Two Text Analytic Approaches. In: Proceedings of the 2nd International Conference on Weblogs and Social Media. 2008 Presented at: AAAI'08; March 30–April 2, 2008; Seattle, Washington URL:https://www.aaai.org/Papers/ICWSM/2008/ICWSM08-020.pdf
32. Chung C, Pennebaker J. The psychological functions of function words. In: Fiedler K, editor. Social Communication: Frontiers of Social Psychology. First Edition. New York: Psychology Press; 2007:343-359.
33. Prieto VM, Matos S, Álvarez M, Cacheda F, Oliveira JL. Twitter: a good place to detect health conditions. PLoS One 2014;9(1):e86191 [FREE Full text] [doi: 10.1371/journal.pone.0086191] [Medline: 24489699]
34. Thackeray R, Burton SH, Giraud-Carrier C, Rollins S, Draper CR. Using Twitter for breast cancer prevention: an analysis of breast cancer awareness month. BMC Cancer 2013 Oct 29;13:508 [FREE Full text] [doi: 10.1186/1471-2407-13-508] [Medline: 24168075]
35. Pavalanathan U, De Choudhury M. Identity Management and Mental Health Discourse in Social Media. In: Proceedings of the 24th International Conference on World Wide Web. 2015 Presented at: WWW'15 Companion; May 18-22, 2015; Florence, Italy p. 315-321. [doi: 10.1145/2740908.2743049]
36. De Choundhury M, Counts S, Horvitz E. Social Media as a Measurement Tool of Depression in Populations. In: Proceedings of the 5th Annual ACM Web Science Conference. 2013 Presented at: WebSci'13; May 2-4, 2013; Paris, France p. 47-56 URL:https://dl.acm.org/citation.cfm?id=2464480 [doi: 10.1145/2464464.2464480]
37. Coppersmith G, Harman C, Dredze M. Measuring Post Traumatic Stress Disorder in Twitter. In: Proceedings of the 8th International Conference on Weblogs and Social Media. 2014 Presented at: AAAI'14; June 1-4, 2014; Ann Arbor, Michigan p. 579-582 URL:https://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/viewFile/8079/8082
38. Birnbaum ML, Ernala SK, Rizvi AF, De Choudhury M, Kane JM. A collaborative approach to identifying social media markers of schizophrenia by employing machine learning and clinical appraisals. J Med Internet Res 2017 Dec 14;19(8):e289 [FREE Full text] [doi: 10.2196/jmir.7956] [Medline: 28807891]
39. Arseniev-Koehler A, Lee H, McCormick T, Moreno MA. #Proana: pro-eating disorder socialization on Twitter. J Adolesc Health 2016;58(6):659-664. [doi: 10.1016/j.jadohealth.2016.02.012] [Medline: 27080731]
40. Statista. 2019. Leading Countries Based on Number of Twitter Users as of January 2019 (in Millions) URL:https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/[WebCite Cache ID 77Ctc6VPr]
41. Twitter Developer. URL:https://developer.twitter.com/en.html [accessed 2018-11-10] [WebCite Cache ID 778Qhq4Gu]
42. American Psychiatric Association. Diagnostic And Statistical Manual Of Mental Disorders, Fifth Edition. Washington, DC: American Psychiatric Publishing; 2013.
43. Natural Language Tool Kit. URL:https://www.nltk.org/api/nltk.tokenize.html [accessed 2018-10-10] [WebCite Cache ID 778Rzo6pe]
44. Padró L, Stanilovsky E. FreeLing 3.0: Towards Wider Multilinguality. In: Proceedings of the Eighth International Conference on Language Resources and Evaluation. 2012 Presented at: LREC'12; May 21-27, 2012; Istanbul, Turkey p. 2473-2479 URL:http://www.lrec-conf.org/proceedings/lrec2012/pdf/430_Paper.pdf
45. Perez-Rosas V, Banea C, Mihalcea R. Learning Sentiment Lexicons in Spanish. In: Proceedings of the Eighth International Conference on Language Resources and Evaluation. 2012 Presented at: LREC'12; May 21-27, 2012; Istanbul, Turkey p. 3077-3081 URL:http://www.lrec-conf.org/proceedings/lrec2012/pdf/1081_Paper.pdf
46. Cruz FL, Troyano JA, Pontes B, Ortega FJ. Building layered, multilingual sentiment lexicons at synset and lemma levels. Expert Syst Appl 2014 Oct;41(13):5984-5994 [FREE Full text] [doi: 10.1016/j.eswa.2014.04.005]

47. Ekman P, Friesen WV, O'Sullivan M, Chan A, Diacoyanni-Tarlatzis I, Heider K, et al. Universals and cultural differences in the judgments of facial expressions of emotion. J Pers Soc Psychol 1987 Oct;53(4):712-717. [doi: 10.1037/0022-3514.53.4.712] [Medline: 3681648]
48. Sidorov G, Miranda-Jiménez S, Viveros-Jiménez F, Gelbukh A, Castro-Sánchez NA, Castillo F, et al. Empirical Study of Opinion Mining in Spanish Tweets. In: Proceedings of the 11th Mexican International Conference on Artificial Intelligence. 2012 Presented at: MICAI'12; October 27-November 4, 2012; San Luis Potosí, Mexico p. 1-4.
49. Nambisan P, Luo Z, Kapoor A, Patrick TB, Cisler RA. Social Media, Big Data, and Public Health Informatics: Ruminating Behavior of Depression Revealed Through Twitter. In: Proceedings of the 2015 48th Hawaii International Conference on System Sciences. 2015 Presented at: HICSS'15; January 5-8, 2015; Kauai, Hawaii p. 2906-2913. [doi: 10.1109/HICSS.2015.351]
50. Cassano P, Fava M. Depression and public health: an overview. J Psychosom Res 2002 Oct;53(4):849-857. [Medline: 12377293]
51. Morris DW, Rush AJ, Jain S, Fava M, Wisniewski SR, Balasubramani GK, et al. Diurnal mood variation in outpatients with major depressive disorder: implications for DSM-V from an analysis of the sequenced treatment alternatives to relieve depression study data. J Clin Psychiatry 2007 Sep;68(9):1339-1347. [doi: 10.4088/JCP.v68n0903] [Medline: 17915971]
52. De Choudhury M, Kiciman E, Dredze M, Coppersmith G, Kumar M. Discovering Shifts to Suicidal Ideation from Mental Health Content in Social Media. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. 2016 Presented at: CHI'16; May 7-12, 2016; San Jose, California p. 2098-2110 URL:http://europepmc.org/abstract/MED/29082385 [doi: 10.1145/2858036.2858207]
53. Bucci W, Freedman N. The language of depression. Bull Menninger Clin 1981 Jul;45(4):334-358. [Medline: 6176285]
54. Rude S, Gortner EM, Pennebaker J. Language use of depressed and depression-vulnerable college students. Cogn Emot 2004;18(8):1121-1133. [doi: 10.1080/02699930441000030]
55. Pyszczynski T, Greenberg J. Self-regulatory perseveration and the depressive self-focusing style: a self-awareness theory of reactive depression. Psychol Bull 1987 Jul;102(1):122-138. [doi: 10.1037/0033-2909.102.1.122] [Medline: 3615702]
56. Pennebaker JW, Mehl MR, Niederhoffer KG. Psychological aspects of natural language use: our words, our selves. Annu Rev Psychol 2003;54:547-577. [doi: 10.1146/annurev.psych.54.101601.145041] [Medline: 12185209]
57. Sonnenschein AR, Hofmann SG, Ziegelmayer T, Lutz WL. Linguistic analysis of patients with mood and anxiety disorders during cognitive behavioral therapy. Cogn Behav Ther 2018;47(4):315-327. [doi: 10.1080/16506073.2017.1419505] [Medline: 29345528]
58. Painuly N, Sharan P, Mattoo SK. Relationship of anger and anger attacks with depression: a brief review. Eur Arch Psychiatry Clin Neurosci 2005 Aug;255(4):215-222. [doi: 10.1007/s00406-004-0539-5] [Medline: 16133740]
59. Neacsiu AD, Rompogren J, Eberle JW, McMahon K. Changes in problematic anger, shame, and disgust in anxious and depressed adults undergoing treatment for emotion dysregulation. Behav Ther 2018;49(3):344-359 [FREE Full text] [doi: 10.1016/j.beth.2017.10.004] [Medline: 29704965]
60. Bernard JD, Baddeley JL, Rodriguez BF, Burke PA. Depression, language, and affect: an examination of the influence of baseline depression and affect induction on language. J Lang Soc Psychol 2016;35(3):317-326. [doi: 10.1177/0261927X15589186]
61. De Choudhury M, Sharma SS, Logar T, Eekhout W, Nielsen RC. Gender and Cross-Cultural Differences in Social Media Disclosures of Mental Illness. In: Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing. 2017 Presented at: CSCW'17; February 25-March 1, 2017; Portland, Oregon p. 353-369. [doi: 10.1145/2998181.2998220]
62. De Choudhury M, De S. Mental Health Discourse on Reddit: Self-Disclosure, Social Support, and Anonymity. In: Proceedings of the 8th International Conference on Weblogs and Social Media. 2014 Presented at: AAAI'14; June 1-4, 2014; Ann Arbor, Michigan p. 71-80 URL:https://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8075/8107
63. Mowery D, Smith H, Cheney T, Stoddard G, Coppersmith G, Bryan C, et al. Understanding depressive symptoms and psychosocial stressors on Twitter: a corpus-based study. J Med Internet Res 2017 Dec 28;19(2):e48 [FREE Full text] [doi: 10.2196/jmir.6895] [Medline: 28246066]
64. Eichstaedt JC, Schwartz HA, Kern ML, Park G, Labarthe DR, Merchant RM, et al. Psychological language on Twitter predicts county-level heart disease mortality. Psychol Sci 2015 Feb;26(2):159-169 [FREE Full text] [doi: 10.1177/0956797614557867] [Medline: 25605707]
65. Yom-Tov E, Johansson-Cox I, Lampos V, Hayward AC. Estimating the secondary attack rate and serial interval of influenza-like illnesses using social media. Influenza Other Respir Viruses 2015 Jul;9(4):191-199 [FREE Full text] [doi: 10.1111/irv.12321] [Medline: 25962320]
66. Paul MJ, Dredze M. Discovering health topics in social media using topic models. PLoS One 2014;9(8):e103408 [FREE Full text] [doi: 10.1371/journal.pone.0103408] [Medline: 25084530]
67. Chary M, Genes N, McKenzie A, Manini AF. Leveraging social networks for toxicovigilance. J Med Toxicol 2013 Jun;9(2):184-191 [FREE Full text] [doi: 10.1007/s13181-013-0299-6] [Medline: 23619711]

68. Seabrook EM, Kern ML, Fulcher BD, Rickard NS. Predicting depression from language-based emotion dynamics: longitudinal analysis of Facebook and Twitter status updates. J Med Internet Res 2018 Dec 8;20(5):e168 [FREE Full text] [doi: 10.2196/jmir.9267] [Medline: 29739736]
69. Reece AG, Reagan AJ, Lix KL, Dodds PS, Danforth CM, Langer EJ. Forecasting the onset and course of mental illness with Twitter data. Sci Rep 2017 Dec 11;7(1):13006 [FREE Full text] [doi: 10.1038/s41598-017-12961-9] [Medline: 29021528]
70. Alvarez-Mon MA, Asunsolo Del Barco A, Lahera G, Quintero J, Ferre F, Pereira-Sanchez V, et al. Increasing interest of mass communication media and the general public in the distribution of Tweets about mental disorders: observational study. J Med Internet Res 2018 Dec 28;20(5):e205 [FREE Full text] [doi: 10.2196/jmir.9582] [Medline: 29807880]

## Abbreviations

**API:** Application Programming Interface
**POS:** Part-of-Speech
**WHO:** World Health Organization

*Edited by G Eysenbach; submitted 29.03.19; peer-reviewed by JL Oliveira, E Gabarron, X Chen, V Pereira-Sanchez; comments to author 23.04.19; revised version received 24.05.19; accepted 24.05.19; published 27.06.19*

*Please cite as:*
*Leis A, Ronzano F, Mayer MA, Furlong LI, Sanz F*
*Detecting Signs of Depression in Tweets in Spanish: Behavioral and Linguistic Analysis*
*J Med Internet Res 2019;21(6):e14199*
*URL: http://www.jmir.org/2019/6/e14199/*
*doi: 10.2196/14199*
*PMID: 31250832*

## 4.3. Evaluating behavioral and linguistic changes during drug treatment for depression: a pairwise study using tweets in Spanish

Depressive disorders are the most common mental illnesses and the leading cause of disability worldwide. Selective serotonin reuptake inhibitors (SSRIs) are the most commonly prescribed drugs for the treatment of depressive disorders. Some people share information about their experiences with antidepressants in social media platforms. The analysis of the messages posted by Twitter users under SSRI treatment can yield useful information on how these antidepressants affect users' behavior. This study aims to compare the behavioral and linguistic characteristics of the tweets posted while users were likely to be under SSRI treatment, in comparison to the tweets posted by the same users when they were less likely to be taking this medication.

Behavioral and linguistic changes have been detected when users with depression are taking antidepressant medication. These features can provide interesting insights for monitoring the evolution of this disease, as well as offering additional information related to treatment adherence. This information may be especially useful in patients who are receiving long-term treatments such as people suffering from depression.

Leis, A., Ronzano, F., Mayer, M. A., Furlong, L. I., & Sanz, F. (2020). Evaluating behavioral and linguistic changes during drug treatment for depression: a pairwise study using tweets in Spanish. *Journal of Medical Internet Research (in press)*.

# Evaluating behavioral and linguistic changes during drug treatment for depression: a pairwise study using tweets in Spanish

Angela Leis, Francesco Ronzano, Miguel Angel Mayer, Laura I. Furlong, Ferran Sanz

Submitted to: Journal of Medical Internet Research
on: June 01, 2020

# Original Manuscript

# Evaluating behavioral and linguistic changes during drug treatment for depression: a pairwise study using tweets in Spanish

## Original Paper

Angela Leis, PsyM; Francesco Ronzano, PhD; Miguel A Mayer, MDH PhD MD; Laura I. Furlong, PhD; Ferran Sanz, Prof Dr

Research Programme on Biomedical Informatics (GRIB), Hospital del Mar Medical Research Institute (IMIM), Department of Experimental and Health Sciences (DCEXS), Universitat Pompeu Fabra, Barcelona, Spain

## Abstract

**Background:** Depressive disorders are the most common mental illnesses and they constitute the leading cause of disability worldwide. Selective serotonin reuptake inhibitors (SSRIs) are the most commonly prescribed drugs for the treatment of depressive disorders. Some people share information about their experiences with antidepressants in social media platforms such as Twitter. The analysis of the messages posted by Twitter users under SSRI treatment can yield useful information on how these antidepressants affect users' behavior.
**Objective:** This study aims to compare the behavioral and linguistic characteristics of the tweets posted while users were likely to be under SSRI treatment, in comparison to the tweets posted by the same users when they were less likely to be taking this medication.
**Methods:** In the first step, the timelines of Twitter users mentioning SSRI antidepressants in their tweets were selected using a list of 128 generic and brand names of SSRIs. In the second step, two datasets of tweets were created, the *in-treatment* dataset (made up of the tweets posted throughout the 30 days after mentioning an SSRI) and the *unknown-treatment* dataset (made up of tweets posted more than 90 days before and more than 90 days after any tweet mentioning an SSRI). For each user, the changes in behavioral and linguistic features between the tweets classified in these two datasets were analyzed. 186 users and their timelines with 668,842 tweets were finally included in the study.
**Results:**
The number of tweets generated per day by the users when they were *in-treatment* was higher than when they were in the *unknown-treatment* period ($P$ = .001). When the users were in treatment, the mean percentage of tweets posted during the daytime (from 8:00 am to midnight) increased with respect to *unknown-treatment* period ($P$ = .002). The number of characters and words per tweet was higher when the users were in treatment ($P$ = .03 and P = .02 respectively). Regarding linguistic features, the percentages of the first and second-person singular pronouns were higher when users were in treatment ($P$ = .008 and $P$ = .004, respectively).
**Conclusions:** Behavioral and linguistic changes have been detected when users with depression are taking antidepressant medication. These features can provide interesting insights for monitoring the evolution of this disease, as well as offering additional information related to treatment adherence. This information may be especially useful in patients who are receiving long-term treatments such as people suffering from depression.

**Keywords:** depression; antidepressant drugs; serotonin uptake inhibitors; mental health; social

media; data mining.

## Introduction

### Background

Depression is one of the most common mental disorders [1]. According to the World Health Organization (WHO) depression affects more than 322 million people of all ages globally, being a leading cause of disability worldwide [2]. The proportion of people with depression went up by around 18% between 2005 and 2015 [3]. This mental disorder constitutes a challenge for society and healthcare systems due to devastating personal and social consequences and the associated economic costs [4-13]. In spite of the high prevalence of depression and the efforts of healthcare services for the improvement of its management, this health condition remains underdiagnosed and undertreated [14].

The pharmacological treatment in the case of moderate and severe forms of depression can improve the quality of life of these patients [4]. There are several types of antidepressant drugs and, among them, the selective serotonin reuptake inhibitors (SSRIs) are currently the most prescribed antidepressants around the world. For instance, according to the Spanish Medicines Agency [15], SSRIs constitutes more than the 70% of all antidepressants prescribed in Spain. They have fewer side effects than other antidepressants [16], show a good risk-benefit ratio [17,18], are safer and better tolerated [19] and exhibit a reduced risk of toxicity in overdose in comparison to tricyclic antidepressants [20]. They are commonly used as first-line treatment for depression [21-23] and are usually prescribed as maintenance therapy to prevent relapse [4,23-26]. SSRIs include the following drugs: fluvoxamine, fluoxetine, paroxetine, sertraline, citalopram and escitalopram [17].

Furthermore, although social media platforms have not been created with health-related purposes in mind [27,28], everyday millions of people publicly share personal health information on social media platforms [29,30]. For this reason, these platforms represent an important source of health information, which is faster and more broadly available than other sources of health information, being unsolicited, spontaneous and up-to-date. Infodemiology approaches have been developed and applied to better understand the dynamics of these platforms when used as a health information source [31-33]. In this context, social media users share health-related information, such as experiences with prescribed drugs [34], cancer patients' sentiments [35], opinions on vaccines [36], or online conversations on epidemic outbreaks [37]. The massive data from social media can be monitored and analyzed by using natural language processing and machine learning technologies, providing new possibilities to better understand users' behavior [30], including automatic identification of early signs of mental disorders [38-40]. In particular, it is usual that people suffering from depression talk about their illness and the drugs they are taking [41-43].

Twitter is a very popular microblogging platform with more than 330 million active users worldwide [44]. Tweets, freely available in almost 90% of users' accounts, provide a huge amount of data that can collected in real-time [28,30,33,45-47]. Twitter users post short messages about facts, feelings and opinions, including health conditions [49].

Mining of drug-related information from Twitter has been applied in the pharmacovigilance field [27,50]. Some pharmacovigilance studies carried out in Twitter studied specific cohorts by identifying users' mentions of drug intake [37,51-53]. Other studies focused on the adverse drug reactions, analyzing users' tweets regarding adverse events and side effects associated to drug use, which were identified by means of generic or brand names [29,47,54,55]. In our previous study [49], we observed that Twitter users who are potentially suffering from depression show particular behavioral and linguistic features in their tweets. These features were related to an increase in their activity during the night, a different style of writing with an increase of the first-person singular pronoun, lower number characters in the tweets, an increase in the frequency of words related to sadness and disgust emotions and more frequent presence of negation words and negative polarity. This information can be used as a complementary tool to detect signals of depression and for monitoring and supporting patients using Twitter.

## Objectives

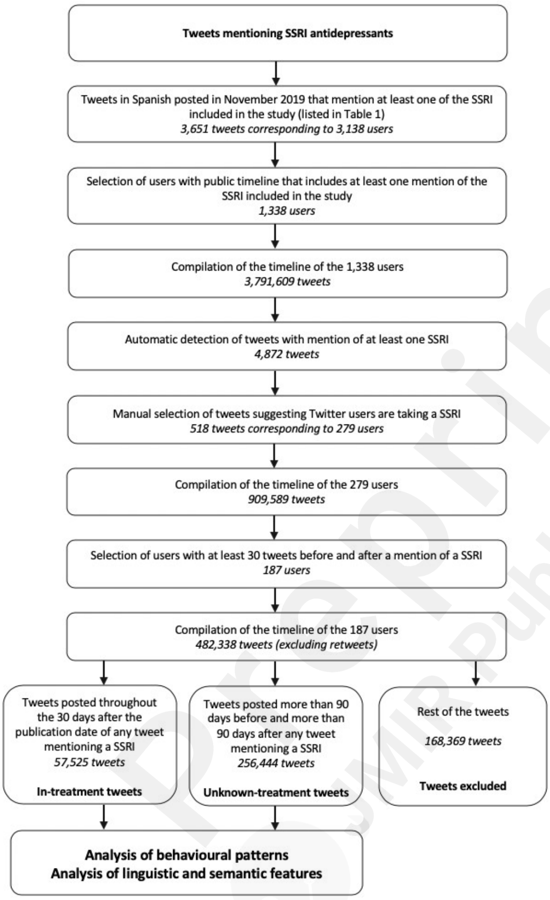
In the present paper, we aim to enrich our previous study [49] by focusing the new one on the analysis of the changes in behavioral and linguistic features of Twitter users in Spanish language, which may be associated to the antidepressant medication these users are taking. It is worth mentioning that users from Spanish-speaking countries are among the most active on Twitter in the world [54]. The study is focused on Twitter users that mention treatment with selective serotonin reuptake inhibitors (SSRIs), which are the most frequently prescribed antidepressants [15]. In particular, this study compares the characteristics of the tweets posted while users were probably taking SSRIs, versus the tweets posted by the same users when they have lower probability of taking this antidepressant medication. This analysis can contribute to better understand how these drugs affect users' mood. Although we found two additional studies describing changes in Twitter users' language in some mental disorders [57,58], to the best of our knowledge, there are no other studies that analyze Twitter posts in Spanish language to detect behavioral and linguistic changes when the users are taking antidepressant medication.

# Methods

## Study Design

This study was designed with the aim of analyzing the behavioral patterns and linguistic features of users that have mentions of selective serotonin reuptake inhibitors (SSRIs) in their Twitter timeline. The study was developed in several steps and focused on tweets written in Spanish. The flow diagram of the study is depicted in Figure 1.

**Figure 1.** Flow diagram of the study process.






As shown in Figure 1 two non-overlapping datasets of tweets from users mentioning treatment with SSRIs were obtained:

1. The *in-treatment tweets dataset*, which was made up of the tweets posted throughout the 30 days after the publication date of any tweet mentioning SSRI intake. We assumed that these tweets were posted while the users had a high probability of being in treatment with an SSRI.

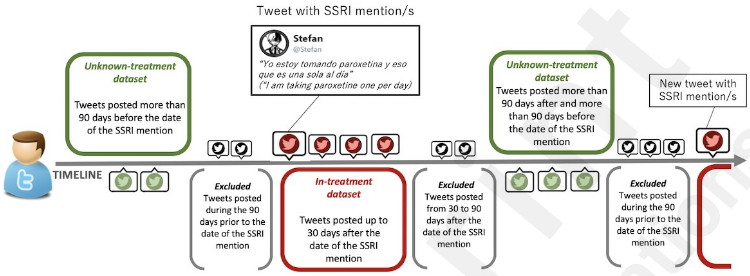
2. The *unknown-treatment tweets dataset*, which was made up of the tweets that were posted more than 90 days before and more than 90 days after the publication date of any tweet mentioning SSRI intake. We assumed that these tweets were posted while users had a lower

probability of being in treatment with an SSRI than in the previous dataset.

These datasets were designed in a way that made possible to carry out intra-subject comparisons since the *in-treatment* tweets and the *unknown-treatment* tweets datasets were obtained from the same Twitter users.

The strategy for the selection of the tweets included in the two datasets is depicted in Figure 2.

**Figure 2.** The *in-treatment* and *unknown-treatment* datasets selection strategy.




## Data Collection and User Selection

The selection of the tweets and their users was based on the filtered real-time streaming support provided by the Twitter API [59]. In the first step, we selected tweets in Spanish that mention any of the SSRIs generic and brand names used around the world. To obtain the generic and brand names, we performed searches on the following databases and resources: DrugBank [60], the Anatomical Therapeutic Chemical Classification System (ATC) and the Defined Daily Dose (DDD) of the World Health Organization [61], Wikipedia [62], and the Database for Pharmacoepidemiological Research in Primary Care (BIFAP) [63]. The list of 135 generic and brand names obtained are shown in Table 1.

**Table 1.** Selective serotonin reuptake inhibitors (SSRIs) used in the study.

| Generic name | Brand names |
|---|---|
| Fluvoxamina (fluvoxamine) | Dumirox, Faverin, Floxyfral, Fluvoxin, Luvox, Uvox |
| Fluoxetina (fluoxetine) | Prozac, Reneuron, Adofen, Luramon, Sarafem |
| Paroxetina (paroxetine) | Seroxat, Motivan, Frosinor, Praxil, Daparox, Xetin, |

| | |
|---|---|
| | Apo-oxpar, Appoxar, Aropax, Aroxat, Aroxat CR, Bectam, Benepax, Casbol, Cebrilin, Deroxat, Hemtrixil, Ixicrol, Loxamine, Meplar, Olane, Optipar, Oxetine, Pamax, ParadiseCR, Paradox, Paraxyle, Parexis, Paroxat, Paroxet, Paxan, Paxera, Paxil, Paxil CR, Pexot, Plasare, Pondera, Posivyl, Psicoasten, Rexetin, Seretran, Sereupin, Tiarix, Tamcere, Traviata, Xerenex, Xetroran |
| Sertralina (sertraline) | Aremis, Besitran, Zoloft, Altisben, Aserin, Altruline, Ariale, Asertral, Atenix, Eleval, Emergen, Dominium, Inosert, Irradial,Sedora, Serolux, Sertex |
| Citalopram (citalopram) | Seropram, Celexa, Akarin , C Pram S , Celapram, Celica, Ciazil, Cilate, Cilift, Cimal, Cipralex, Cipram, Cipramil, Cipraned, Cinapen, Cipramine Ciprotan Citabax, Citaxin, Citalec, Citalex, Citalo, Citalopram, Citol, Citox, Citrol, Citta, Dalsan, Denyl, Elopram, Estar, Humorup, Humorap, Oropram, Opra, Pram, Pramcit, Procimax, Recital, Sepram, Szetalo, Talam, Temperax, Vodelax, Zentius, Zetalo, Cipratal, Zylotex |
| Escitalopram (escitalopram) | Cipralex, Diprex, Esertia, Essential, Heipram |

The following seven brand names of medicines have been excluded due to their semantic ambiguity: Essential, Motivan, Estar, Traviata, Pondera, Recital, Emergen. These commercial names are, at the same time, very common words used with different meanings in Spanish as we verified after reviewing a random sample of 200 tweets with mentions of these words. The number of tweets excluded because of their semantic ambiguity was 21,104. In a manual checking of random samples of them (n=200), the mentions of SSRIs when using these words were 0% (0/200) in some cases, such as Motivan and Estar, and 0.5% (1/200) in Recital. Eventually, the final list of words included 128 generic and brand names of SSRIs.

Using the aforementioned 128 SSRI names, we collected 3,651 tweets in Spanish posted during November 2019 with occurrences of the words listed in Table 1. These tweets were posted by 3,138 different Twitter users and mentioned 33 different words of the list. The frequencies of these 33 words are shown in Table 2.

**Table 2.** Frequencies of SSRI names mentioned in Spanish tweets during November 2019.

| **SSRI mentions** | **Frequency** |
|---|---|
| Prozac | 998 |
| Fluoxetina | 756 |
| Sertralina | 542 |
| Escitalopram | 248 |
| Citta | 210 |
| Citalo | 109 |

| | |
|---|---|
| Paroxetina | 69 |
| Pram | 49 |
| Fluvoxamina | 40 |
| Citalopram | 33 |
| Seroxat | 22 |
| Eleval | 21 |
| Lexapro | 20 |

| | |
|---|---|
| Opra | 18 |
| Casbol | 14 |
| Ariale | 11 |
| Zoloft | 9 |
| Altruline | 9 |
| Paxil | 7 |
| Akarin | 7 |

| | |
|---|---|
| Heipram | 4 |
| Aremis | 4 |
| Cimal | 3 |
| Tiarix | 2 |
| Seretran | 2 |
| Dominium | 2 |
| Citox | 2 |

| | |
|---|---|
| Atenix | 2 |
| Aserin | 2 |
| Talam | 1 |
| Dalsam | 1 |
| Celexa | 1 |

In a second step, we crawled the public Twitter timelines of the 3,138 users (until 3,200 most recent tweets for each user). Given that retweets are not useful to analyze the linguistic behavior of a particular user, the third step consisted of excluding the retweets and checking if the remaining tweets of each timeline included the mention of at least one SSRI. 1,800 users were excluded by this filter, leaving a total of 1,338 Twitter users. We obtained 3,791,609 tweets after compiling the timelines from these 1,338 users. From these timelines, 4,872 tweets mentioning at least one of the SSRIs from the list were automatically detected. These 4,872 tweets were independently reviewed by two experts, a psychologist and a family physician, both with clinical experience. These experts manually selected the tweets that suggested that the user who posted the tweet was taking an SSRI on the date of posting. Examples of these tweets are shown in Textbox 1.

**Textbox 1**. Examples of positive and negative tweets suggesting about taking an SSRI.

Positive examples:

- *"Eso de tener sueños raros debido a la fluoxetina se está saliendo de control."* *("Having odd dreams due to fluoxetine is getting out of control")*

- *"Yo tomo sertralina, como me lo receta el doctor y aún así a veces siento que el mundo donde estoy no es para mi. Ese susto esa angustia esas ganas de correr es algo que sólo el que lo padece lo entiende"*
  *("I take sertraline as my doctor prescribes it to me and, even so, sometimes I feel that the world I'm living in is not for me. This fear this anxiety this desire to run out is something that only who suffers from it can understand")*

Negative examples:

- *"Ella debería tomar prozac, como Tic Tac"*
  *("She should take prozac, like Tic Tac" (a candy brand))*

- *"La Paroxetina es un medicamento que pertenece a la familia de los antidepresivos inhibidores de la recaptación de la serotonina ¡Conoce más sobre él!"*
  *("Paroxetine is a drug that belongs to the antidepressant family of serotonin reuptake inhibitors. Find out more about it!")*

The agreement between reviewers was 93.3% (4,537/4,872) with a Cohen's Kappa score equal to 0.68, indicating that there was a substantial agreement between raters. The reviewers discussed and reached a consensus on the classification of the 335 tweets they classified differently. Finally, we obtained a total number of 518 tweets with one or more SSRI mentions, suggesting that the users who posted these tweets were taking an SSRI at the moment of posting. These tweets corresponded to 279 different users. Therefore, these users had two characteristics: first, the tweets of their timeline included at least one mention of SSRIs and, second, the text of tweets mentioning SSRIs suggested that the user was taking the antidepressant. In addition, we analyzed the tweets posted by each user that belonged to the two datasets (*in-treatment* and *unknown-treatment*, see Fig. 1) by trying different minimum numbers of tweets per dataset (10, 30, 60 and 100 tweets) in order to include a user in the study. 10 tweets contained few information in terms of number of words or posting characteristics. In the case of 60 and 100 tweets, the number of users included dropped dramatically. For this reason, we applied a requirement of a minimum of 30 tweets in both *in-treatment* and *unknown-treatment* datasets to keep the balance between the number of tweets and the number of users to be included in the study. After applying this requirement, 187 users were finally included in the study. The complete timelines of these users were compiled totalizing 668,842 tweets, which were reduced to 482,338 once retweets were removed. Out of these, 168,369 more tweets were excluded because they were posted on dates located outside the periods that qualified a tweet for being included in the *in-treatment* or the *unknown-treatment* datasets. Finally, 57,525 tweets were included in the *in-treatment dataset* and 256,444 in the *unknown-treatment dataset*.

## Data Analysis

The two datasets of tweets, *in-treatment* and *unknown-treatment*, were compared in order to determine the existence of behavioral and linguistic differences between the tweets generated by the users in each period. The features that were analyzed are listed in Table 3.

**Table 3**. Features of the tweets analyzed.

| **Features** | **Analyses performed** |
|---|---|

| | |
|---|---|
| Distribution over time | Tweets per hour<br>Tweets during daytime vs. night<br>Tweets per day<br>Tweets during weekdays vs. weekend |
| Length | Number of characters<br>Number of words |
| Part of Speech (POS) | Number of words by grammatical categories (part-of speech tags) |
| Emotions analysis | Frequencies of emotion types |
| Negations | Frequencies of negation words |
| Polarity | Polarity of tweets on the basis of Spanish Sentiment Lexicon |

Paired data statistical significance tests (paired t-tests) were carried out whenever possible. Benjamini-Hochberg false discovery rate was applied as multiple testing correction analysis [64]. The p-values provided incorporate it.

The textual content of each tweet was analyzed using the same methodology and tools of our previous study [49]. The textual content of each tweet was analyzed by means of the following steps:

- Tokenization performed based on a customized Twitter tokenizer included in the Natural Language Toolkit [65].
- Part-of-Speech (POS) tagging performed by means of the Freeling Natural Language Processing tool in order to analyze the usage patterns of grammatical categories, such as verbs, nouns, pronouns, adverbs and adjectives, in the text of tweets [66].
- Identification of negations performed by building upon a customized list of Spanish negation expressions, such as *nada* (nothing), *nadie* (nobody), *no* (no), *nunca* (never), and alike.

- Identification of positive and negative words inside the text of each tweet using the Spanish Sentiment Lexicon [67].
- Identification of words and expressions associated to emotions, such as happiness, anger, fear, disgust, surprise and sadness [68] by using the Spanish Emotion Lexicon [69].

The statistical analyses were carried out using Python 3.7, the Tweepy, Scipy, NLTK libraries, and R version 3.6.2 (Development Core Team) including the R ´psych´ package 1.9.12.31. All the aforementioned software tools are publicly available.

## Ethical Approval

The protocol used in this study was reviewed and approved by the Ethics Committee of Parc Salut Mar (approval number 2017/7234/1).

## Results

### *Distribution Over Time*

Several types of distribution-over-time analysis were performed in order to study the potential influence of being in *in-treatment* periods in comparison to *unknown-treatment* ones. The tweet hours were adjusted by the users' time zone.

The mean of the time period analyzed of all the users was 28.2 months (SD 24.7) and the mean of the total number of tweets analyzed corresponding to *in-treatment* periods was 307.6 (SD 336.0) and 1,371.4 (SD 748.2) in the case of *unknown-treatment* periods. The mean number of tweets per day generated by users when they were *in-treatment* periods was 11.44 (SD 10.05); such number dropped to 9.07 (SD 7.21) in the *unknown-treatment* dataset with a mean difference of 2.37 (SD 9.72) between periods, which shows statistically significant differences between both datasets ($t_{186}$=3.33; *P* < .001).

Regarding the tweets posted between 8:00 am and midnight (daytime), the mean percentage of tweets posted during daytime was 64.30% (SD 14.83) when the users were *in-treatment* periods; this percentage fell to 61.78% (SD 13.69) during the *unknown-treatment* periods, with a mean percentage difference of 2.52% (SD 11.81), which implies statistically significant differences ($t_{186}$=3.07; *P* = .004).

The mean number of tweets generated during the weekdays (from Monday to Friday) was 12.28 (SD 11.05) when *in-treatment* and 9.33 (SD 6.70) in the *unknown-treatment* periods, with a mean difference of 2.95 (SD 10.23) and statistically significant differences between both datasets ($t_{186}$=3.93; *P* < .001). In relation to the mean number of tweets generated during the weekends (Saturday and Sunday), it was 9.35 (SD 9.31) in the *in-treatment* period and 8.41 (SD 9.82) in the *unknown-treatment* period, with a mean difference of 0.94 (SD 10.92) that implies statistically significant differences between both datasets ($t_{186}$=1.18; *P* = .23). Concerning the percentage of tweets posted throughout the weekdays, the mean percentage over these days was 75.95% (SD 9.17) when the users were *in-treatment*; such percentage went down to 74.40% (SD 5.31) in *unknown-treatment* period, with a mean percentage difference of 1.56% (SD 8.9) that implies statistically significant differences between the two periods ($t_{186}$=2.39; *P* = .018).

### *Length*

The average number of characters per tweet was 88.03 (SD 30.74) and 85.19 (SD 28.82) in the *in-treatment* and *unknown-treatment* datasets respectively, with a mean difference of 2.84 (SD 17.70) and statistically significant differences between both periods ($t_{186}$=2.19; P = .03). As for the number of words per tweet, the mean was 15.68 (SD 5.75) in the *in-treatment* dataset and 15.09 (SD 5.20) in the *unknown-treatment* dataset, with a mean difference of 0.59 (SD 3.54) and statistically significant differences ($t_{186}$=2,28; P = .02).

### *Links and mentions to other users*

The mean percentages of tweets that include at least one link were 23.10 (SD 16.16) and 23.27 (SD 15.29) in the *in-treatment* and *unknown-treatment* datasets respectively, with a mean difference of -0.17 (SD 10.94), a difference that is not statistically significant ($t_{186}$= -0.23; P = .82). Regarding the mean percentages of tweets that include at least one mention to another Twitter user were 45.79 (SD 24.77) and 43.52 (SD 24.71) in the *in-treatment* and *unknown-treatment* datasets respectively, with a mean difference of 2.27 (SD 12.13), which is statistically significant ($t_{186}$= 2.56; P = .011).

### *Part-of-Speech*

As for the analysis of the number of words by grammatic category (i.e. Part-of-Speech) in each tweet, we also compared the *in-treatment* and *unknown-treatment* datasets. The mean percentage of words per grammatical category over the total number of words in each dataset is shown in Table 4. We considered the most relevant lexical POS such as verbs, nouns, pronouns, adverbs and adjectives, excluding conjunctions, interjections, punctuations, determiners, adpositions, numbers or dates.

**Table 4**. Percentages of Part-of-Speech (POS) words comparing *in-treatment* and *unknown-treatment* datasets.

| Type of POS | Mean of percentages *in-treatment* | Mean of percentages *unknown-treatment* | Mean of differences | SD of differences | Paired t-test | p-value |
|---|---|---|---|---|---|---|
| Verbs | 18.50 | 18.20 | 0.3 | 1.28 | 3.15 | .002 |
| Nouns | 19.50 | 19.94 | -0.44 | 2.57 | -2.35 | .020 |
| Pronouns | 9.19 | 8.93 | 0.26 | 1.33 | 2.61 | .010 |
| Adverbs | 6.42 | 6.36 | 0.06 | 0.84 | 0.97 | .335 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Adjectives | 6.05 | 6.21 | -0.16 | 0.95 | -2.34 | .020 |

Regarding the different types of pronouns, the mean percentages of personal pronouns in each dataset are shown and compared in Table 5.

**Table 5**. Mean percentages of personal pronouns comparing *in-treatment* and *unknown-treatment* datasets

| Personal pronouns | Mean of percentages *in-treatment* | Mean of percentages *unknown-treatment* | Mean of differences | SD of differences | Paired t-test | p-value |
|---|---|---|---|---|---|---|
| 1st person singular | 49.50 | 47.80 | 1.7 | 8.68 | 2.67 | .008 |
| 2nd person singular | 14.77 | 16.07 | -1.3 | 6.17 | -2.88 | .004 |
| 3th person singular | 22.13 | 22.86 | -0.73 | 5.79 | -1.72 | .08 |
| 1st person plural | 3.44 | 3.43 | 0.01 | 3.43 | 0.04 | .96 |
| 2nd person plural | 0.998 | 0.9998 | 0 | 1.22 | -0.01 | .98 |
| 3th person plural | 5.60 | 5.39 | 0.21 | 3.68 | 0.77 | .44 |

## *Emotion Analysis*

The mean percentages of the different emotions obtained using the Spanish Sentiment Lexicon on the tweets posted in the two periods, are shown in Table 6.

**Table 6**. Mean percentages of different emotions comparing *in-treatment* and *unknown-treatment* datasets

| Emotion | Mean of percentages *in-treatment* | Mean of percentages *unknown-treatment* | Mean of differences | SD of differences | Paired t-test | p-value |
|---|---|---|---|---|---|---|
| Happiness | 26.93 | 25.94 | 0.99 | 5.82 | 2.32 | .02 |
| Sadness | 10.01 | 9.76 | 0.25 | 4.20 | 0.81 | .41 |
| Fear | 3.20 | 3.02 | 0.18 | 1.94 | 1.23 | .21 |
| Anger | 5.52 | 5.20 | 0.32 | 2.71 | 1.62 | .11 |
| Disgust | 3.11 | 3.06 | 0.05 | 1.97 | 0.38 | .69 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Surprise | 5.59 | 5.06 | -1.47 | 2.42 | 2.98 | .003 |

### *Negation Analysis*

The mean percentages of tweets, among all users, that included one or more negation words were 27.66 (SD 10.54) and 26.59 (SD 9.87) for the *in-treatment* and *unknown-treatment* datasets respectively, with a mean difference of 1.07 (SD 6.99), which is statistically significant ($t_{186}$= 2.10; *P* = .037).

### *Polarity Analysis*

As for the polarity of tweets, the percentage of occurrences, among all users, of one or more positive words inside the text of each tweet over total was 15.13% (SD 6.56) in the *in-treatment* dataset and 14.50% (SD 5,43) in the *unknown-treatment* dataset, with a mean percentage difference of 0.63% (SD 5.22) ($t_{186}$= 1.66; *P* = .09). The percentage of one or more negative words in each tweet over total was 7.97% the *in-treatment* dataset (SD 4.40) and 7.54% (SD 3.52) in the *unknown-treatment* dataset, with a mean percentage difference of 0.43% (SD 3.58) ($t_{186}$= 1.64; *P* = .10). No statistically significant differences were detected in this analysis.

## Discussion

### *Principal Findings*

Social media platforms in general and Twitter in particular, may provide useful information on how patients respond when they receive a pharmacological treatment, as has been shown in several studies in which social media can be used as a complementary source of pharmacovigilance and monitoring [34,70]. In this study we analyzed the tweets of users who mentioned they were taking antidepressant drugs, in particular SSRIs, with the aim of detecting behavioral changes when they are more likely to be in treatment in comparison to periods in which they are less likely to be in treatment (*"in-treatment"* vs. *"unknown-treatment"* periods).

The results of the present study show that, in general, Twitter users significantly increased their activity of posting tweets during the *in-treatment* periods. This increase was more pronounced during weekdays in comparison to weekends. We also observed a significantly greater proportion of tweets posted during the daytime during the *in-treatment* periods. These results are consistent with the results of our previous paper [49] in which we observed that the control group without signs of depression showed more tweet posting activity than the group of users with signs of depression, especially during the daytime and the weekdays. These results are also consistent with another paper that described the behavior in social media of people with self-reported depression [41], as well as with a study on the diurnal mood variation of patients suffering from major depressive disorder [71]. In summary, we can state that when considering the tweet posting activity, the behavior of individuals suffering from depression becomes closer to the general population when they are in treatment with SSRI.

Likewise, the average number of characters and words per tweet were significantly higher when the Twitter users were in treatment with SSRI, a finding that again points towards an increase on the activity of these treated users. In addition, the increase in the number of mentions per tweet can reflect a greater interest in interacting with other people. All these changes may be due to some improvement in their anhedonic symptomatology because of the medication.

Regarding the linguistic analysis, we observed quantitatively slight changes between the *in-treatment* and the *unknown-treatment* periods, although in some cases they are statistically significant. These slight findings are not easily interpretable. In general, given that the style of writing of people suffering from depression is characterized by the self-focus attention, which is associated to negative emotional states and psychological distancing in order to connect with others [72], we can conclude that when the studied subjects were in treatment they improved some traits related to their posting activity as previously mentioned but, at the same time, their language maintained the features of people suffering from depression without a clear influence of the medication.

Emotions is another important aspect that characterizes people suffering from depression and they were consequently analyzed. When the users were in treatment, they showed small but statistically significant improvements in the happiness and surprise emotions, but not in sadness or other emotions (i.e. anger, fear and disgust). As for the number of negations, they slightly increased the use of this these types of words during the *in-treatment* period. However, the polarity analysis did not show differences between both periods.

The increased activity observed on Twitter when the users are likely to be in treatment with SSRIs, can be linked to an improved emotional status in their happiness and surprise emotions. These changes are consistent with our previous observations on mood states of Twitter users without depression compared to depressive ones [49]. However, the traits that are related to the language, as indicated by the POS analysis and the use of negations, maintained a similar profile to that of subjects with depression, independently of the pharmacological treatment detected. These results denote that users with depression who are taking SSRIs show some mood improvements while receiving antidepressant treatment, but at the same time maintain an altered language pattern, which may be indicative of incomplete recovery.

On the basis of our statistically significant results, we may state that Twitter timelines can be used as complementary tool to monitor subjects in order to detect adherence to treatment, which is an important problem in this kind of patients. Adherence to treatment is essential for disease remission [73-76]. According to some studies, it is common that patients suffering from depression do not maintain the duration of antidepressant treatment that is clinically recommended [4,18,77]. In summary, the follow-up of behavioral and language changes in users' Twitter timelines can be useful for monitoring the evolution of the depressive symptoms and the effect of the treatments.

### *Limitations and Future Directions*

This type of studies in general, and this one in particular, presents some limitations. For instance, we considered tweets written in Spanish and from public Twitter users' timelines and these users may be not representative of the general population or from people suffering from depression [33,49,78,79]. Some studies have shown that the Twitter users are often urban people with high levels of education, and they are generally younger than the general population [33,49,78,80,81]. We should also take into account that SSRIs are used in different types of depressive disorders and in other mental conditions. Moreover, we have no information about if these drugs were taken in the context of a prescribed medical treatment or as a result of an unappropriated self-medication decision.

Another limitation may be related with the fact that Twitter users who share their personal drug intake may use words or expressions not included in the list of drug names employed in this study for streaming tweets, even though we tried to be exhaustive in the list of names used. Twitter texts are informal and limited by the number of characters and it is common the inclusion of abbreviations, errors or slang language [33,45]. All these issues can make difficult the automatic extraction of drug mentions and link them to a formal lexicon [28,30,50,53,55]. Unlike clinical records that could be linked to domain resources, the lack of lay vocabularies related to health concepts and terminologies hinder the processing of social media texts [55]. In addition, the results obtained may depend on the particular drugs selected for the study [33], as well as on the periods of time set up for classifying the tweets into the *in-treatment* and in *unknown-treatment* datasets. On the basis of the strategy applied for defining the groups of tweets to be compared (tweets generated just after mentions to SSRI intake vs. tweets generated in periods far from any mention to the SSRI intake), there is some chance of misclassification (it is likely that not all the tweets in first group have been generated by users under actual SSRI treatment, and it is probable that some tweets of the second group have been generated by users under SSRI treatment).

Furthermore, we must take into account that data from social media posts use to contain irrelevant information. Although the proportion of useful information for the specific research purpose can be quite limited, it constitutes a useful starting point [28,30,51,53]. In this scenario the human curation of tweets is a necessary step in this kind of analyses [34]. Even so, due to the different nuances that a tweet can involve, it is not easy to detect real drug intakes or first-hand experiences [24,46,52].

## Conclusions

Social media can be used to monitor the health status of people and, in particular, to detect symptoms or features related to diseases or health conditions by means of the analysis of the users' behavior and language on social media platforms. Moreover, the detection of changes in symptoms or other features when patients are taking medications, can provide interesting insights for monitoring pharmacological treatments, as well as for following up the evolution of the disease, detecting side-effects or providing information related to the treatment adherence. The changes in some features, such as the decrease of the activity on Twitter or the length of tweets, the increase of self-focus through the use of the first-person singular pronoun, and changes in the happiness and surprise emotions could be used as complementary tools to detect the worsening of the psychological status of users suffering from depression, as well as to perceive the lack of adherence to treatment. This information may be especially useful in patients suffering from chronic diseases who are receiving long-term treatments, as is the case of mental disorders in general and depression in particular. However, it is not possible to determine the specific reasons why the individuals change their behavior and language on social media platforms in the framework of a disease and its treatment without performing a clinical assessment. Overall, this study shows the relevance of monitoring the behavioral and linguistic changes in the tweets of persons taking antidepressants. These changes are likely to be influenced by the diverse stages of the disease and the therapeutic effects of the treatment that these Twitter users are receiving, opening a new line of research to better understand and follow up depression through social media.

## Acknowledgements

We received support from the Agency for Management of University and Research Grants in Catalonia (Spain) for the incorporation of new research personnel (FI2016) and from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement number 802750 (FAIRplus) with the support of the European Union's Horizon 2020 research and innovation programme and EFPIA Companies.

The Research Programme on Biomedical Informatics (GRIB) is a member of the Spanish National Bioinformatics Institute (INB), funded by ISCIII and FEDER (PRB2-ISCIII), and it is supported by grant PT17/0009/0014. The DCEXS is a 'Unidad de Excelencia María de Maeztu', funded by the MINECO [MDM-2014-0370]. Funding for open access charge: Agència de Gestió d'Ajuts Universitaris i de Recerca Generalitat de Catalunya [2017 SGR 00519].

The Database for Pharmacoepidemiological Research in Primary Care. Spanish Agency for Medicines and Health Products (AEMPS) of the Ministry of Health and Consumer Affairs and Social Welfare of the Government of Spain was used to obtain useful information about the prescription frequency of antidepressants.

## Conflicts of Interest

None declared.

## Abbreviations

**API:** Application Programming Interface
**POS:** Part-of-Speech
**SSRIs:** Selective serotonin reuptake inhibitors
**WHO:** World Health Organization
**BIFAP:** Database for Pharmacoepidemiological Research in Primary Care (BIFAP)

## References

1. World Health Organization. 2019. Depression: Key Facts. URL: https://www.who.int/news-room/fact-sheets/detail/depression [accessed 2020-01-09]
2. World Health Organization. 2017. Depression and Other Common Mental Disorders: Global Health Estimates URL: http://apps.who.int/iris/bitstream/10665/254610/1/WHO-MSD-MER-2017.2-eng.pdf?ua=1 [accessed 2020-01-09]
3. Ferrari AJ, Charlson FJ, Norman RE, Patten SB, Freedman G, Murray CJ, et al. Burden of depressive disorders by country, sex, age, and year: findings from the global burden of disease study 2010. *PLoS Med.* 2013 Nov;10(11):e1001547. PMID:24223526.
4. *Position statement on antidepressants and depression. PS04/19*. London, United Kingdom: Royal College of Psychiatrist; May 2019.
5. Marcus M, Yasamy MT, van Ommeren M, Chisholm D, Saxena S, WHO Department of Mental Health and Substance Abuse. World Health Organization. 2012. Depression: a global public health concern URL: http://www.who.int/mental_health/management/depression/who_paper_depression_wfmh_2012.pdf
6. Global Burden of Disease Study 2013 Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: a systematic analysis for the global burden of disease study 2013. *Lancet.* 2015 Aug 22;386(9995):743-800. PMID:26063472.

7. Trautmann S, Rehm J, Wittchen HU. The economic costs of mental disorders: do our societies react appropriately to the burden of mental disorders? *EMBO Rep.* 2016;17(9):1245-1249. PMID:27491723.
8. Patel V, Chisholm D, Parikh R, Charlson FJ, Degenhardt L, Dua T, et al. Global priorities for addressing the burden of mental, neurological,substance use disorders. In: Patel V, Chisholm D, Dua T, Laxminarayan R, Medina-Mora ME, Vos T, editors. Disease Control Priorities: Mental, Neurological, and Substance Use Disorders, Third Edition (Volume 4). Washington, DC: The World Bank; 2016:1-27.
9. Whiteford HA, Degenhardt L, Rehm J, Baxter AJ, Ferrari AJ, Erskine HE, et al. Global burden of disease attributable to mental and substance use disorders: findings from the global burden of disease study 2010. *Lancet.* 2013 Nov 9;382(9904):1575-1586. PMID:23993280.
10. Wongkoblap A, Vadillo MA, Curcin V. Researching mental health disorders in the era of social media: systematic review. *J Med Internet Res.* 2017 Dec 29;19(6):e228. PMID:28663166.
11. Working Group of the Clinical Practice Guideline on the Management of Depression in adults. *Clinical practice guideline on the management of depression in adults*. Galicia, Spain: Ministry of Health, Social Services and Equality. Galician Agency for Health Technology Assessment, avalia-t, Ministry of Health, Social Services and Equality; 2014. https://portal.guiasalud.es/wp-content/uploads/2018/12/GPC_534_Depresion_Adulto_Avaliat_compl_en.pdf
12. Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Med*. 2006 Nov;3(11):e442. PMID:17132052.
13. Vigo D, Thornicroft G, Atun R. Estimating the true global burden of mental illness. *Lancet Psychiatry*. 2016 Feb;3(2):171-178. PMID:26851330.
14. Kraus C, Kadriu B, Lanzenberger R, Zarate Jr CA, Kasper S. Prognosis and improved outcomes in major depression: a review. *Trans Psychiatry*. 2019 Apr 3;9(1):1-7.
15. Utilización de medicamentos antidepresivos en España durante el periodo 2000-2013. Informe de utilización de medicamentos U/AD/V1/14012015. Madrid: Ministerio de Sanidad, Servicios Sociales e Igualdad. Agencia Española de Medicamentos y Productos Sanitarios (AEMPS); 2015.
16. Luo Y, Kataoka Y, Ostinelli EG, Cipriani A, Furukawa TA. National prescription patterns of antidepressants in the treatment of adults with major depression in the US between 1996 and 2015: a population representative survey based analysis. *Front Psychiatry*. 2020 Feb 14;11:35. doi: 10.3389/fpsyt.2020.00035. PMID: 32116850.
17. Fasipe OJ. The emergence of new antidepressants for clinical use: Agomelatine paradox versus other novel agents. *IBRO Rep*. 2019 Jan, 9(6):95-110. PMID: 31211282.
18. *Depression in adults: recognition and management. Clinical guideline [CG90]*. London, United Kingdom: National Institute for Health and Care Excellence (NICE); Updated: April 2018. https://www.nice.org.uk/guidance/cg90/resources/depression-in-adults-recognition-and-management-pdf-975742638037
19. Gelenberg AJ, Freeman MP, Markowitz JC, Rosenbaum JF, Thase ME, Trivedi MH, Van Rhoads RS, et al. Practice guideline for the treatment of patients with major depressive disorder. 3rd ed. Washington, DC: American Psychiatric Association; 2010.
20. Lane R, Baldwin D, Preskorn S. The SSRIs: advantages, disadvantages and differences. J Psychopharmacol. 1995;9(suppl 2),163–178. PMID:22297235.

21. Pundiak TM, Case BG, Peselow ED, et al. Discontinuation of maintenance selective serotonin reuptake inhibitor monotherapy after 5 years of stable response: a naturalistic study. *J Clin Psychiatry*. 2008; 69: 1811–1817. PMID:19026252.
22. Emslie GJ, Mayes TL, Ruberu M. Continuation and maintenance therapy of early-onset major depressive disorder. *Paediatr Drugs*. 2005; 7: 203–217. PMID:16117558.
23. Garnock-Jones KP, McCormack PL. Escitalopram: a review of its use in the management of major depressive disorder in adults. *CNS Drugs*. 2010; 24: 769–796. PMID:20806989.
24. Clevenger, SS, Devvrat M, Jonathan D, Brigitte V, Waguih WI. The Role of selective serotonin reuptake inhibitors in preventing relapse of major depressive disorder. *The Adv Psychopharmacol*. 2018; 8(1):49–58. PMID:29344343.
25. Sim K, Lau W, Sim J, Sum M, Baldessarini R. (2015) Prevention of relapse and recurrence in adults with major depressive disorder: Systematic Review and Meta-Analyses of Controlled Trials. *Int J Neuropsychopharmaco*l. 2015 Jul 7;19(2): pyv076. PMID:26152228.
26. Peselow ED, Tobia G, Karamians R, et al. Prophylactic efficacy of fluoxetine, escitalopram, sertraline, paroxetine, and concomitant psychotherapy in major depressive disorder: outcome after long-term follow-up. *Psychiatry Res*. 2015; 225:680-686. PMID:25496869.
27. Salathé M. Digital Pharmacovigilance and disease surveillance. *J Infect Dis*. 2016 Dec 1;214(suppl 4):399-403. PMID:28830106.
28. Sarker A, Ginn R, Nikfarjam A, O'Connor K, Smith K, Jayaraman S, Upadhaya T, Gonzalez G. Utilizing social media data for pharmacovigilance: a review. J *Biomed Inform*. 2015 Apr 1;54:202-12. PMID:25720841.
29. Adrover C, Bodnar T, Huang Z, Telenti A, Salathé M. Identifying adverse effects of HIV drug treatment and associated sentiments using twitter. *JMIR Public Health Surveill*. 2015;1(2):e7. Published 2015 Jul 27. doi:10.2196/publichealth.4488. PMID:27227141.
30. Nikfarjam A, Sarker A, O'Connor K, Ginn R, Gonzalez G. Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features. *J Am Med Inform Assoc*. 2015 May;22(3):671-681.PMID:25755127.
31. Salathé M. Digital epidemiology: what is it, and where is it going? *Life Sci Soc Policy*. 2018 Jan 4;14(1):1. doi: 10.1186/s40504-017-0065-7. PMID:29302758.
32. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the internet. *J Med Internet Res*. 2009 Mar 27;11(1):e11. PMID:19329408.
33. Lardon J, Bellet F, Aboukhamis R, et al. Evaluating Twitter as a complementary data source for pharmacovigilance. *Expert Opin Drug Saf*. 2018 Aug 3;17(8):763-74. PMID:29991282.
34. Freifeld CC, Brownstein JS, Menone CM, Bao W, Filice R, Kass-Hout T, Dasgupta N. Digital drug safety surveillance: monitoring pharmaceutical products in twitter. *Drug Saf*. 2014 May 1;37(5):343-50. PMID:24777653. https://link.springer.com/article/10.1007%2Fs40264-014-0155-x
35. Crannell, WC, Clark, E, Jones, C, James, TA, Moore, J. A pattern-matched Twitter analysis of US cancer-patient sentiments. *J Surg Res*. 2016;*206*(2):536-542. PMID: 27523257
36. Surian D, Nguyen DQ, Kennedy G, Johnson M, Coiera E, Dunn AG. Characterizing Twitter discussions about HPV vaccines using topic modeling and community detection. *J Med Internet Res.* 2016;18(8):e232. doi:10.2196/jmir.6045. PMID:27573910.
37. Chen E, Lerman K, Ferrara E. Tracking social media siscourse about the COVID-19 pandemic: development of a public coronavirus Twitter data set. *JMIR Public Health Surveill*. 2020;6(2):e19273. doi:10.2196/19273. PMID: 32427106.
38. Reece, AG, Reagan, AJ, Lix, KLM, Danforth, CM, Dodds, PS, Langer, EJ. Forecasting the

onset and course of mental illness with Twitter data. *Sci Rep.* 2017;7(1):1-11. doi:10.1038/s41598-017-12961-9
39. Park M, Cha C, Cha M. Depressive moods of users portrayed in Twitter. In: Proceedings of the ACM SIGKDD Workshop on Health Informatics. 2012 Presented at: HI-KDD'12; August 12-16, 2012; Beijing, China p. 1-8.
40. Conway M, O'Connor D. Social media, big data, and mental health: current advances and ethical implications. *Curr Opin Psychol*. 2016 Jun;9:77-82. doi: 10.1016/j.copsyc.2016.01.004. PMID: 27042689.
41. De Choudhury M, Gamon M, Counts S, Horvitz E. Predicting Depression via Social Media. In Proceedings of the Seventh International Conference on Weblogs and Social Media. 2013 Presented at: AAAI'13; July 8-11, 2013; Cambridge, MA p. 128-138.
42. Cavazos-Rehg PA, Krauss MJ, Sowles S, Connolly S, Rosas C, Bharadwaj M, et al. A content analysis of depression-related Tweets. *Comput Human Behav*. 2016;54:351-357. doi: 10.1016/j.chb.2015.08.023 PMID: 26392678
43. Nguyen T, O'Dea B, Larsen M, Phung D, Venkatesh S, Christensen H. Using linguistic and topic analysis to classify sub-groups of online depression communities. *Multimed Tools Appl*. 2015 Dec 21;76(8):10653-10676. doi: 10.1007/s11042-015-3128-x
44. Statista. 2018. Number of Monthly Active Twitter Users Worldwide From 1st Quarter 2010 to 2nd Quarter 2018 (in Millions). URL: https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/ [accessed 2019-02-03]
45. Audeh B, Calvier FE, Bellet F, Beyens MN, Pariente A, Lillo-Le Louet A, Bousquet C. Pharmacology and social media: potentials and biases of web forums for drug mention analysis—case study of France. *Health Informatics J*. 2019 Sep 30:1460458219865128. PMID:31566468.
46. Alvaro N, Conway M, Doan S, Lofi C, Overington J, Collier N. Crowdsourcing Twitter annotations to identify first-hand experiences of prescription drug use. *J Biomed Inform*. 2015 Dec 1;58:280-7. PMID:26556646.
47. Pierce CE, Bouri K, Pamer C, Proestel S, Rodriguez HW, Van Le H, Freifeld CC, Brownstein JS, Walderhaug M, Edwards IR, Dasgupta N. Evaluation of Facebook and Twitter monitoring to detect safety signals for medical products: an analysis of recent FDA safety alerts. *Drug Saf*. 2017 Apr 1;40(4):317-31. PMID:28044249.
48. Bian J, Topaloglu U, Yu F. Towards large-scale twitter mining for drug-related Adverse Events. *SHB 2012*. 2012 October 29, 25-32. PMID:28967001. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5362648/
49. Leis A, Ronzano F, Mayer MA, Furlong LI, Sanz F. Detecting Signs of Depression in tweets in spanish: behavioral and linguistic analysis. *J Med Internet Res*. 2019;21(6):e14199. PMID:31250832.
50. O'Connor K, Pimpalkhute P, Nikfarjam A, Ginn R, Smith KL, Gonzalez G. Pharmacovigilance on twitter? Mining tweets for adverse drug reactions. *AMIA Annu Symp Proc*. 2014 Nov 14;2014:924-933. PMID: 25954400.
51. Mahata D, Friedrichs J, Ratn R, Jiang J. Did you take the pill? Detecting personal intake of medicine from Twitter. *IEEE Intell Syst*. 2018 Jul/Aug;33(4):87-95. doi:10.1109/MIS.2018.043741326
52. Klein A, Sarker A, Rouhizadeh M, O'Connor K, Gonzalez G. Detecting personal medication intake in Twitter: an annotated corpus and baseline classification system. In Proceedings of the 16th Biomedical Natural Language Processing (BioNLP). 2017. Vancouver, BC, Canada, Aug 4:136-142. https://www.aclweb.org/anthology/W17-2316.pdf
53. Kiritchenko S, Mohammad SM, Morin J, de Bruijn B. NRC-Canada at SMM4H shared task: classifying tweets mentioning adverse drug reactions and medication intake. In Proceedings

of the 2nd Social Media Mining for Health Applications Workshop co-located with the American Medical Informatics Association Annual Symposium (AMIA). 2017. Washington DC, Nov 4, 2017:1-11. http://ceur-ws.org/Vol-1996/paper1.pdf
54. Nikfarjam, Azadeh Ransohoff JD, Callahan A, et al. Early detection of adverse drug reactions in social health networks: a natural language processing pipeline for signal detection. *JMIR Public Health Surveill. 2019 Jun 3;*5(2):e11264. 3. PMID:31162134.
55. Segura-Bedmar I, Martínez P, Revert R, Moreno-Schneider J. Exploring spanish health social media for detecting drug effects. *BMC Med Inform Decis Mak*. 2015;15 Suppl 2:S6. PMID:26100267.
56. Statista. 2019. Leading Countries Based on Number of Twitter Users as of April 2020 (in Millions) URL: https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/
57. Liu J, Weitzman ER, Chunara R. Assessing sehavioral stages from social media data. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 2017 Presented at: CSCW 2017; February 25-March 1, 2017; Portland, Oregon p. 1320-1333.
58. Koustuv S, Benjamin S, Torous J, Abrahao B, Kıcıman E, De Choudhury M. A social media study on the effects of psychiatric medication use. In *Proceedings of the thirteenth International Conference on Web and Social Media.* 2019. Presented at: ICWSM 2019; Juny 11-14; Munich, Germany p 440-451.
59. Twitter Developer. URL: https://developer.twitter.com/en.html. [accessed 2019-12-16].
60. Wishart DS, Knox C, Guo AC, et al. DrugBank: a knowledgebase for drugs, drug actions and drug targets. Nucleic Acids Res. 2008;36 (Database issue):D901–D906. PMID:18048412.
61. WHO Collaborating Centre for Drug Statistics Methodology, Norwegian Institute of Public Health. *The Anatomical Therapeutics Chemical Classification System (ATC)*. https://www.whocc.no/. Last updated December 16, 2019. Accessed May 6, 2019.
62. *Selective serotonin reuptake inhibitor*. (2019, May 7). Wikipedia. Retrieved May 7, 2019 from https://en.wikipedia.org/wiki/Selective_serotonin_reuptake_inhibitor
63. Spanish Agency for Medicines and Health Products (AEMPS), Ministry of Health and Consumer Affairs and Social Welfare. *Multiregional primary care database of Spanish population (BIFAP)*. http://www.bifap.org/. Accessed May 6, 2019.
64. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple hypothesis testing. *J R Stat Soc B.* 1995;57(1):289-300
65. Natural Language Tool Kit. URL: https://www.nltk.org/api/nltk.tokenize.html [accessed 2019-12-10]
66. Padró L, Stanilovsky E. FreeLing 3.0: Towards Wider Multilinguality. In: Proceedings of the Eighth International Conference on Language Resources and Evaluation. 2012 Presented at: LREC'12; May 21-27, 2012; Istanbul, Turkey p. 2473-2479 URL: http://www.lrec-conf.org/proceedings/lrec2012/pdf/430_Paper.pdf
67. Perez-Rosas V, Banea C, Mihalcea R. Learning Sentiment Lexicons in Spanish. In: Proceedings of the Eighth International Conference on Language Resources and Evaluation. 2012 Presented at: LREC'12; May 21-27, 2012; Istanbul, Turkey p. 3077-3081. URL: http://www.lrec-conf.org/proceedings/lrec2012/pdf/1081_Paper.pdf
68. Ekman P, Friesen WV, O'Sullivan M, Chan A, Diacoyanni-Tarlatzis I, Heider K, et al. Universals and cultural differences in the judgments of facial expressions of emotion. *J Pers Soc Psychol*. 1987 Oct;53(4):712-717. PMID:3681648.
69. Sidorov G, Miranda-Jiménez S, Viveros-Jiménez F, Gelbukh A, Castro-Sánchez NA, Castillo F, et al. Empirical Study of Opinion Mining in Spanish Tweets. In: Proceedings of the 11th Mexican International Conference on Artificial Intelligence. 2012 Presented at: MICAI'12;

October 27-November 4, 2012; San Luis Potosí, Mexico p. 1-4.
70. Carbonell P, Mayer MA, Bravo À. Exploring brand-name drug mentions on Twitter for pharmacovigilance. *Stud Health Technol Inform*. 2015;210:55-9. PMID: 25991101.
71. Morris DW, Rush AJ, Jain S, Fava M, Wisniewski SR, Balasubramani GK, et al. Diurnal mood variation in outpatients with major depressive disorder: implications for DSM-V from an analysis of the sequenced treatment alternatives to relieve depression study data. *J Clin Psychiatry.* 2007 Sep;68(9):1339-1347. PMID:17915971.
72. Pennebaker JW, Mehl MR, Niederhoffer KG. Psychological aspects of natural language use: our words, our selves. *Annu Rev Psychol*. 2003;54:547-77. PMID:12185209.
73. Anghelescu IG, Kohnen R, Szegedi A, et al. Comparison of hypericum extract WS 5570 and paroxetine in ongoing treatment after recovery from an episode of moderate to severe depression: results from a randomized multicenter study. *Pharmacopsychiatry.* 2006; 39: 213–219.PMID:17124643.
74. Martin-Vazquez MJ. Adherence to antidepressants: a review of the literature. *Neuropsychiatry*. 2016;6(5):236-41. PMID:22808448.
75. Mitchell AJ. Depressed patients and treatment adherence. *Lancet*. 2006 Jun 24;367(9528):2041-2043. PMID:16798371.
76. De las Cuevas, C., Peñate, W. & Sanz, E.J. Risk factors for non-adherence to antidepressant treatment in patients with mood disorders. *Eur J Clin Pharmacol*. 2014;70, 89–98. PMID:24013851.
77. De Choudhury M, De S. Mental Health Discourse on Reddit: Self-Disclosure, Social Support, and Anonymity. In: Proceedings of the 8th International Conference on Weblogs and Social Media. 2014 Presented at: AAAI'14; June 1-4, 2014; Ann Arbor, Michigan p. 71-80 URL: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8075/8107
78. Sadah SA, Shahbazi M, Wiley MT, Hristidis V. A Study of the Demographics of Web-Based Health-Related Social Media Users. *J Med Internet Res*. 2015;17(8):e194.
79. Eichstaedt JC, Schwartz HA, Kern ML, Park G, Labarthe DR, Merchant RM, et al. Psychological language on Twitter predicts county-level heart disease mortality. *Psychol Sci.* 2015 Feb;26(2):159-169. PMID:25605707.
80. Yom-Tov E, Johansson-Cox I, Lampos V, Hayward AC. Estimating the secondary attack rate and serial interval of influenza-like illnesses using social media. *Influenza Other Respir Viruses.* 2015 Jul;9(4):191-199. PMID:25962320.
81. Paul MJ, Dredze M. Discovering health topics in social media using topic models. *PLoS One.* 2014;9(8):e103408. PMID:25084530.

# 5. DISCUSSION

*"The truth is rarely pure and never simple"*

Oscar Wilde

## 5.1. Overview

Given the heterogeneous nature of this depressive disorders and the variability of symptoms among the individuals suffering from them, the prevention, diagnosis, and treatment of the depressive disorders is a complicated task, which result in a great number of depressive disorder cases that are undetected or inadequately treated (Cassano & Fava, 2002; Nambisan et al., 2015; Nguyen et al., 2015). For these reasons and taking into account that people diagnosed with depression are increasing worldwide, new strategies for detecting and monitoring this disease would be very useful, including the use of social media platforms as a source of health information.

As have been shown in previous studies on Twitter (Cavazos et al., 2016; De Choudhury, Counts, Horvitz, 2013; De Choudhury, Gamon, Counts et al., 2013; Mowery et al. 2017; Nambisan et al., 2015; Park et al., 2012; Reece et al.; 2017; Wilson & Valstar, 2014).) and our own ones carried out in the present thesis, users share personal information including health data. In addition, Twitter may provide useful information on how patients respond when they receive a pharmacological treatment, and for this reason it can be used as a complementary source of pharmacovigilance and monitoring (Carbonell et al., 2015; Freifeld et al., 2014). In the present thesis, we have analyzed the behavioral and linguistic patterns of tweets in the Spanish language that suggest signs of depression with the aim of supporting Twitter as a complementary health source of information to detect and monitor depressive patients. These technologies can provide diverse advantages such as performing a large-scale and remote assessment, helping professionals the management of these patients and giving the opportunity to monitor and support them continuously (Morales, 2018).

## 5.2. Big Data: Challenges and perspectives in the context of social media and psychiatric conditions

Big Data is a concept that describes a new generation of technologies that allow us to extract valuable information from massive data volumes and types. According to the International Medical Informatics Association (IMIA) Working Group on Data Mining and Big Data Analytics, Big Data can be defined as data whose scale, diversity, and complexity require new architecture, techniques, algorithms, and analytics to manage it and extract value and hidden knowledge from it (Bellazi, 2014; Mayer, 2016).

In the last few years an increasing interest in obtaining health information from sources beyond electronic health records has found Big Data an ally. As it was previously mentioned, social media platforms are one of the most important sources of massive data that contain useful health information. Social media can be considered a new complementary channel of health information, and although it is generally of much lower quality, the huge amount of data that is available in these environments can facilitate the development of several activities such as health promotion, disease surveillance and pharmacovigilance.

There are several issues related to the use of social media as a source of health information from the Big Data analytics point of view. From the technical side, we should take into account that the information is very heterogeneous and requires a previous preparation process to allows its fruitful analysis. In addition, the quality and completeness of the data may widely vary, which can affect its interpretation and the knowledge derived from the analysis. Besides, it is critical to make use of the appropriate methodologies and analytic tools to manage and interpret this type of data. Additionally, when mining health personal Big Data, potential conflicts related to data privacy and protection may arise, taking into account that personal health data is considered a very sensitive type of information. Legal and ethical issues constitute important aspects in the context of Big Data research. In this scenario, several guidelines and recommendations of good practices regarding health-related data have been drawn up (Australian Medical Association Council, 2020; British Medical Association, 2018; World Medical Association, 2017), although there is not yet a general agreement among health professionals on how some situations that arise when social media data are analyzed should be managed (Brown, 2010; Thompson et al., 2011).

Our results show that the tweets of depressive users present different behavioral patterns and linguistic features, even when their tweets are not related to depression. The behavioral patterns of depressive users indicate that they are more active during the night in comparison to the rest of the Tweeter users, suggesting that they increase their posting activity in this period of time due to insomnia, which is one of the most frequent symptoms of depression, as other studies pointed out (De Choudhury, Gamon, Counts et al., 2013). In addition, the daily mood changes, such as the morning and evening worsening that are characteristic common in several forms of depression, could justify the lower activity of the depressive users during day hours (Morris et al., 2007). Moreover, the depressive users presented a more regular activity throughout the week, with an increased activity during the weekend in comparison with the rest of the users. This trend may be related to the lowered social activity of the people suffering from depressive disorders, having a reduced participation in social leisure activities during the weekend and spending more time at home, sharing their feelings and thoughts on social media platforms (Cavazos-Rehg, 2016). On the other hand, when considering the tweet posting activity, the behavior of individuals suffering from depression in treatment with SSRI becomes closer to that of the general population.

On the other hand, our findings suggest that the language of people suffering from depression is characterized by a different style of writing that some authors describe as poorly structured, indicating less interest in what surrounds them, people, objects, or things (De

Choudhury et al., 2016). Consistent with many previous studies (Bucci & Freedman, 1981; De Choudhury, Gamon, Counts et al., 2013; Pavalanathan & De Choudhury, 2015; Pyszczynski & Greenberg, 1987; Rude et al., 2004; Tausczik & Pennebaker, 2009), the use of first-person singular pronoun is more frequent among depressive users, demonstrating the attention to self-focus associated with the negative emotional states of depression and the reduced attentional resources, highlighting the psychological distancing to connect with others (Pennebaker & Mehl, 2003). Based on the characteristic of tweets, it is possible to detect potentials depressed users on Twitter (De Choudhury, Counts & Horvitz, 2013).

The negatively focused emotion language, which is typical in depressive patients and feelings is detected by means of the frequent presence of negation words in the tweets of depressed users, as previously pointed out by other authors (Bernard et al., 2016; Ramirez-Esparza et al., 2008; Rude et al., 2004; Pyszczynski & Greenberg, 1987). Therefore, language can be used as a measure of different individual features since people's word choice is stable over time and consistent across topics or context and, consequently, language style appears to be a useful predictor of depression (De Choudhury, Gamon, Counts et al., 2013; Pavalanathan & De Choudhury, 2015). Moreover, the tweets posted by people suffering from depression contain a smaller number of characters and this might be related to reduced interest and poorer language.

In many mental health conditions and, particularly, when people are suffering from depression, emotions are one of the key aspects. The self-focus state that characterizes depression is associated with negative emotions (Chung & Pennebaker, 2007; Pennebaker et al., 2003; Sonnenschein et al., 2018). Regarding negative emotions, we observed an increase in the frequency of words related to sadness, which was also observed in other studies (Park et al., 2012; Pavalanathan & De Choudhury 2015; Sonnenschein et al., 2018); anger, due to this patients tend to feel irritable, wronged, or angry at the world (Cavazos-Rehg et al., 2016; Pavalanathan & De Choudhury, 2015; Painuly et al., 2005; Park et al., 2012) and disgust, an emotion that is known to be associated with the depressive disorders (Neacsiu et al., 2018). People suffering from depression tend to focus more on negative aspects of their life, and this is reflected in the negative polarity of the words used in their tweets posted, as previous studies have shown (De Choudhury, Gamon, Counts et al., 2013; Pavalanathan & De Choudhury, 2015).

In summary, Twitter timelines can be used as a complementary tool to detect people suffering from depression all well as to monitor the evolution of the depressive symptoms, as well as the effect and adherence to treatment, which is essential for disease remission and it is an important problem in this type of patients (Anghelescu et al., 2006; Martin-Vazquez, 2016; De las Cuevas et al., 2014).

## 5.3. Future prospects

There are diverse opportunities and challenges regarding the analysis of social media data with health-related purposes. First, additional research might be carried out to examine specific depression types and determine if there are social media features that can contribute to classifying users or tweets to the different diagnosis of depression (Reece et al., 2017). Second, the possible relationship between depression and seasonality could be of interest for future studies in the context of monitoring Twitter activity (Alverez-Mon et al., 2018). Moreover, the development of studies in which clinical records' information and social media data are linked allowing us to analyse the existence of significant relationships, in which patients potentially modify the features of posting on social media depending on the clinical progress they experience. In this sense, the detection of some changes on social media can predict the onset of particular symptoms (e.g., suicidal ideation).

As we have detected in this thesis, the studies based on the analysis of social media platforms such as Twitter, presents some limitations. For instance, we considered tweets written in Spanish and from public Twitter users' timelines and these users may be not representative of the general population or from people suffering from depression (Eichstaedt et al., 2015; Lardon et al., 2018; Pierce et al., 2017; Sadah et al., 2015). Moreover, some studies have shown that the Twitter users are often urban people with high levels of education, and they are generally younger than the general population

(Lardon et al., 2018; Paul & Drezde; 2014; Sadah et al., 2015; YomTov et al., 2015). On the other hand, without clinically assessing the people declaring depression in their Tweeter profile, there is no way to verify if the diagnosis is genuine or if they suffer from another mental disorder.

It is important to highlight that the approach and methodologies used in this thesis were multidisciplinary, including the collaborative participation of mental health professionals and bioinformaticians, essential collaboration to achieve the objectives proposed. These collaborations are essential to face some of mental disorders' main' challenges by using digital technology (Birnbaum et al., 2017). By means of interdisciplinary collaborations, it is possible to develop digital apps (Huguet et al., 2016) and web services (Justicia et al., 2017) providing personalized alerts and psychosocial support in the mental health domain, making the most of Mobile Health. The introduction of these methods and tools for the automatic detection of signals of depression on social media platforms, has the potential of being used as complementary instruments for the assessment of these patients, assisting healthcare professionals in the detection and monitoring of mental health disorders. Likewise, it is critical to develop studies in which clinical information and social media data are linked in order to get a complete health picture of patients, inside and outside clinical environments.

The results of the studies carried out in this thesis will contribute to the growing body of scientific literature that analyzes the messages posted on social media using languages other than English, which is very important taking into account that the great majority of the world population does not use English as primary language for the verbal communication. The thorough list of words in Spanish used by patients suffering from depression in clinical environments and that was scored by health professionals, is publicly available in GitHub at https://cutt.ly/ThTHrkG. Finally, the dataset of curated tweets that was created in one of the studies described in the results section can be used as a source of information for performing machine learning procedures. This curated and anonymized dataset is publicly available in Kaggle at https://cutt.ly/RhTGEdU. These procedures can be used to train mathematical algorithms in order to perform semi-supervised experiments that will allow computational systems to learn automatic identification and classification of massive datasets, with the aim of being used to detect tweets and Tweeter users with depressive features and behaviors.

# 6. CONCLUSIONS

*"You only live once, but if you do it right,*
*once is enough"*
Mae West

1. We have proved, for the first time, that several behavioral and linguistic features of the tweets in Spanish can be used as a complementary tool to detect signals of depression of their authors, corroborating and extending the findings obtained by studies carried out on English tweets.

2. We have demonstrated that Twitter users who are potentially suffering from depression significantly modify the traits of their language, independently from the fact that the tweets are related or not related to the direct expression of depression.

3. We have shown that the analysis of social media posts related to depression requires to consider not only their linguistic features, but also behavioral patterns, such as the activity over time or the length of tweets.

4. We have shown that changes in the language and other features of posts on social media when users are taking medications, can provide interesting new insights for monitoring the pharmacological treatments of depression.

5. Two publicly available resources have been created, on the one hand, a thorough list of words used by patients suffering from depression, and on the other hand, a curated dataset of tweets indicative of depression, which can be used as a basis for developing new studies and strategies for the analysis of depression on Twitter in the Spanish language.

6. We have shown that the introduction of new strategies and methods for the automatic detection of signals of depression on social media platforms, such as Twitter, has the potential of being used as a complementary tool for the follow-up of depression patients and their pharmacological treatment.

# 7. APPENDIXES

## Appendix 1: Publications included in the results section of the present thesis

- Leis, A., Ronzano, F., Mayer, M. A., Furlong, L. I., & Sanz, F. (2019). Detecting signs of depression in tweets in Spanish: behavioral and linguistic analysis. *Journal of Medical Internet Research*, *21*(6), e14199.

- Leis, A., Mayer, M. A., Ronzano, F., Torrens, M., Castillo, C., Furlong, L. I., & Sanz, F. (2020). Clinical-Based and Expert Selection of Terms Related to Depression for Twitter Streaming and Language Analysis. *Studies in Health Technology and Informatics*, *270*, 921-925.

- Leis, A., Ronzano, F., Mayer, M. A., Furlong, L. I., & Sanz, F. (2020). Evaluating behavioral and linguistic changes during drug treatment for depression: a pairwise study using tweets in Spanish. *Journal of Medical Internet Research (in press)*

# Appendix 2: Other publications of the author

- Mayer, M. A., Gutiérrez-Sacristán, A., Leis, A., De La Peña, S., Sanz, F., & Furlong, L.I. (2017). Using electronic health records to assess depression and cancer comorbidities. *Studies in Health Technology and Informatics, 235, 236-240*.

- Gutiérrez-Sacristán, A., Bravo, À., Portero-Tresserra, M., Valverde, O., Armario, A., Blanco-Gandía, M. C., Farré, A., Fernández-Ibarrondo, L., Fonseca, F., Giraldo, J., Leis, A., Mané, A., Mayer, M. A., Montagud-Romero, S., Nadal, R., Ortiz, J., Pavon, F. J., Perez, E.J., Rodríguez-Arias, M., …& Furlong, L. I. (2017). Text mining and expert curation to develop a database on psychiatric diseases and their genes. *Database* (Oxford), 2017, bax043.

- Mayer, M. A., Fernández-Luque, L., & Leis, A. (2016). Big Data For Health Through Social Media. In S. Syed-Abdul, E. Gabarron & A. Lau, *Participatory Health Through Social Media* (pp. 67-82). Academic Press. Elsevier Inc.

# Appendix 3: Author's contributions in conferences and workshops

- Leis, A., Ronzano, F., Mayer, M. A., Castillo, C., Torrens, M., Furlong, L. I., & Sanz, F. (2019, April). Detecting depression on social media for supporting patients' management [Poster]. In *European Psychiatry* (Vol. 56, pp. S1047, E-PP0435). France: Elsevier France-Editions Scientifiques Medicales Elsevier.

- Leis, A., Ronzano, F., & Mayer, M. A. (2018). *El Virus de la Inmunodeficiencia Humana en Twitter* [Presentation]. III Jornades Internacionals Comunicació i Salut #ParlemdeVIH? Institut de la Comunicació, Universitat Autònoma de Barcelona, Barcelona, Spain.

- Gutiérrez-Sacristán, A., Bravo, À., Portero-Tresserra, M., Valverde, O., Armario, A., Blanco-Gandía, M. C., Farré, A., Fernández-Ibarrondo, L., Fonseca, F., Giraldo, J., Leis, A., Mané, A., Mayer, M. A., Montagud-Romero, S., Nadal, R., Ortiz, J., Pavon, F. J., Perez, E. J., Rodríguez-Arias, M., … & Furlong, L. I. (2017, June). *PsyGeNET: a knowledge resource on psychiatric diseases and their genes* [Poster presentation]. Translational Bioinformatics Wellcome Genome Campus Hinxton, Cambridge, United Kingdom.

# Appendix 4: Spanish stopwords

un, una, unas, unos, uno, sobre, todo, también, tras, otro, algún, alguno, alguna, algunos, algunas, ser, es, soy, eres, somos, sois, estoy, esta, estamos, estáis, están, cómo, en, para, atrás, porque, por qué, estado, estaba, ante, antes, siendo, ambos, pero, por, poder, puede, puedo, podemos, podéis, pueden, fui, fue, fuimos, fueron, hacer, hago, hace, hacemos, hacéis, hacen, cada, fin, incluso, primero desde, conseguir, consigo, consigue, consigues, conseguimos, consiguen, ir, voy, va, vamos, vais, van, vaya, bueno, ha, tener, tengo, tiene, tenemos, tenéis, tienen, el, la, lo, las, los, su, aquí, mío, tuyo, ellos, ellas, nos, nosotros, vosotros, vosotras, dentro, solo, solamente, saber, sabes, sabe, sabemos, sabéis, saben, último, largo, bastante, haces, muchos, aquellos, aquellas, sus, entonces, tiempo, verdad, verdadero, verdadera, cierto, ciertos, cierta, ciertas, intentar, intento, intenta, intentas, intentamos, intentais, intentan, dos, bajo, arriba, encima, usar, uso, usas, usa, usamos, usais, usan, emplear, empleo, empleas, emplean, ampliamos, empleais, valor, muy, era, eras, éramos, eran, modo, bien, cual, cuando, donde, mientras, quien, con, entre, sin, trabajo, trabajar, trabajas, trabaja, trabajamos, trabajáis, trabajan, podria, podrias, podríamos, podrían, podriais, yo, aquel, y, a, e, i, o, u.

# Appendix 5: e-mail sent to psychiatrics and family physicians inviting them to participate in the questionnaire on words used by depressed patients in clinical environments

**In Catalan**

Benvolguts companys,

Ens posem en contacte amb vosaltres per demanar la vostra col·laboració en l'estudi "EMRedesS: Malalties mentals a les Xarxes Socials", que s'està duent a terme des del Research Programme on Biomedical Informatics (GRIB) del Departament de Ciències Experimentals i de la Salut (*DCEXS*-UPF) en col·laboració amb el Hospital del Mar, i que ha estat aprovat pel Comitè Ètic de Recerca Clínica del Parc Salut Mar.

L'objectiu d'aquest estudi és caracteritzar, mitjançant l'anàlisi dels missatges en Twitter i/o Facebook en espanyol, la presència de subjectes amb trets o símptomes de depressió.

Per a això ens seria de gran utilitat si poguéssiu avaluar el llistat de paraules inclòs en l'arxiu adjunt a aquest missatge, en funció de la freqüència d'ús d'aquestes paraules per part dels pacients a l'hora de descriure i verbalitzar el seu trastorn depressiu, segons la vostra experiència. En una escala d'1 a 5, es tractaria de puntuar aquestes paraules en funció de la freqüència d'ús:

1 Mai
2 Rarament
3 Ocasionalment
4 Freqüentment
5 Molt freqüentment

D'aquesta forma, per exemple si puntuem amb un 1 una paraula determinada, estarem indicant que en la nostra experiència, mai ha estat utilitzada pels nostres pacients.

Així mateix, us agrairíem si a més poguéssiu incloure algunes frases o expressions que els pacients utilitzen en la consulta a l'hora de descriure els seus símptomes o estat d'ànim en relació a la depressió. Això pot ocórrer tant en paraules freqüentment utilitzades com en aquelles que mai solen utilitzar-se. Per exemple: “anhedònia” es podria puntuar amb un 1 perquè mai s'utilitza, però en canvi es poden expressar mitjançant frases com: “tot m'és igual”, “gens m'interessa” o “gens em fa il·lusió”.

Una vegada finalitzada l'avaluació, us agrairíem que enviéssiu l'arxiu a Angela Leis, email: angela.leis@upf.edu, si es possible abans del 30 de setembre. No dubteu a contactar amb ella per a qualsevol dubte que us pot sorgir.

Moltes gràcies per la vostra col·laboració.

**In Spanish**

Apreciados compañeros,

Nos ponemos en contacto con vosotros para pedir vuestra colaboración en el estudio “EMRedesS: Enfermedades mentales en las Redes Sociales”, que se está llevando a cabo desde el Research Programme on Biomedical Informatics (GRIB) del Departamento de Ciencias Experimentales y de la Salud (*DCEXS*-UPF) en colaboración con el Hospital del Mar, y que ha sido aprobado por el Comité Ético de Investigación Clínica del Parc Salut Mar.

El objetivo de este estudio es caracterizar, mediante el análisis de los mensajes en Twitter y/o Facebook en español, la presencia de sujetos con rasgos o síntomas de depresión.

Para ello nos sería de gran utilidad si pudierais evaluar el listado de palabras incluido en el archivo adjunto a este mensaje en función de la frecuencia de uso de dichas palabras por parte de los pacientes a la hora de describir y verbalizar su trastorno depresivo, según vuestra experiencia. En una escala de 1 a 5, se trataría de puntuar dichas palabras en función de la frecuencia de uso:

1 Nunca
2 Raramente
3 Ocasionalmente
4 Frecuentemente
5 Muy frecuentemente

De esta forma, por ejemplo, si puntuamos con un 1 una palabra determinada, estaremos indicando que en nuestra experiencia, nunca ha sido utilizada por nuestros pacientes.

Asimismo, os agradeceríamos si además pudieras incluir algunas frases o expresiones que los pacientes utilizan en la consulta a la hora de describir sus síntomas o estado de ánimo en relación con la depresión. Esto puede ocurrir tanto en palabras frecuentemente utilizadas como en aquellas que nunca suelen utilizarse. Por ejemplo: "anhedonia" se podría puntuar con un 1 porque nunca se utiliza, pero en cambio se pueden expresar mediante frases como: "todo me da igual", "nada me interesa" o "nada me hace ilusión".

Una vez que finalizada vuestra evaluación, os agradeceríamos que enviarais el archivo a Angela Leis, email: angela.leis@upf.edu, si es posible antes del 30 de septiembre. No dudéis en contactar con ella para cualquier duda que os puede surgir.

Muchas gracias por vuestra colaboración.

## Appendix 6: Scores obtained in the questionnaire on words used by depressed patients in clinical settings

The list of 255 words and their scores obtained in the questionnaire on words used by depressed patients in clinical settings are listed below and are available at GitHub for research purposes:
https://github.com/angelaleism/WordsDepression

| Words | P[a]1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 | P9 | P10 | P11 | P12 | P13 | P14 | P15 | F[b]1 | F2 | F3 | F4 | F5 | Score |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Deprimido/a | 5 | 5 | 5 | 5 | 4 | 5 | 5 | 4 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 4 | 5 | 5 | 5 | 5 | 97 |
| Triste | 5 | 5 | 5 | 5 | 5 | 5 | 4 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 4 | 4 | 5 | 97 |
| Tristeza | 5 | 5 | 5 | 5 | 5 | 5 | 4 | 5 | 5 | 5 | 5 | 5 | 5 | 4 | 5 | 5 | 5 | 4 | 4 | 5 | 96 |
| Desanimado/a | 5 | 5 | 5 | 5 | 4 | 5 | 5 | 4 | 5 | 5 | 5 | 5 | 4 | 5 | 5 | 4 | 4 | 4 | 5 | 5 | 94 |
| Depresión | 5 | 4 | 5 | 5 | 4 | 5 | 5 | 4 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 3 | 5 | 4 | 3 | 5 | 92 |
| Depresivo/a | 5 | 5 | 5 | 5 | 4 | 5 | 5 | 3 | 4 | 5 | 5 | 5 | 5 | 5 | 5 | 4 | 5 | 4 | 3 | 5 | 92 |
| Ansiedad | 4 | 5 | 5 | 4 | 4 | 5 | 4 | 4 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 3 | 5 | 4 | 4 | 5 | 91 |
| Cansado/a | 4 | 5 | 5 | 4 | 4 | 5 | 5 | 4 | 5 | 4 | 5 | 4 | 5 | 5 | 5 | 4 | 5 | 4 | 4 | 5 | 91 |
| Lloro | 4 | 5 | 5 | 4 | 4 | 5 | 5 | 5 | 5 | 5 | 4 | 3 | 5 | 5 | 4 | 5 | 5 | 4 | 4 | 4 | 90 |
| Agobio | 3 | 5 | 5 | 5 | 3 | 5 | 5 | 4 | 5 | 5 | 5 | 5 | 4 | 4 | 4 | 5 | 5 | 3 | 4 | 5 | 89 |
| Ansioso/a | 4 | 5 | 4 | 3 | 4 | 5 | 4 | 4 | 5 | 5 | 5 | 5 | 5 | 5 | 4 | 3 | 5 | 4 | 5 | 5 | 89 |
| Insomnio | 5 | 4 | 5 | 5 | 5 | 5 | 5 | 4 | 5 | 5 | 3 | 5 | 5 | 5 | 5 | 4 | 5 | 2 | 3 | 4 | 89 |
| Nervioso/a | 4 | 5 | 5 | 5 | 4 | 5 | 2 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 3 | 4 | 3 | 4 | 5 | 89 |
| Agobiado/a | 3 | 5 | 5 | 4 | 3 | 5 | 5 | 4 | 5 | 5 | 5 | 5 | 4 | 4 | 4 | 5 | 4 | 3 | 4 | 5 | 87 |
| Angustiado/a | 4 | 5 | 5 | 3 | 3 | 5 | 5 | 4 | 5 | 5 | 4 | 4 | 5 | 5 | 3 | 4 | 5 | 4 | 4 | 5 | 87 |
| Angustia | 4 | 5 | 4 | 3 | 3 | 5 | 5 | 4 | 4 | 5 | 4 | 5 | 5 | 5 | 5 | 4 | 5 | 4 | 3 | 5 | 87 |
| Agotado/a | 3 | 5 | 5 | 3 | 4 | 5 | 4 | 4 | 5 | 4 | 5 | 4 | 4 | 5 | 5 | 4 | 4 | 4 | 4 | 5 | 86 |
| Decaído/a | 4 | 5 | 4 | 4 | 5 | 4 | 4 | 4 | 4 | 5 | 5 | 5 | 4 | 5 | 3 | 4 | 5 | 3 | 4 | 5 | 86 |

[a] Psychiatrist
[b] Family physician

| Words | | P2 | P3 | P4 | P5 | P6 | P7 | P8 | P9 | P10 | P11 | P12 | P13 | P14 | P15 | F[b]1 | F2 | F3 | F4 | F5 | Score |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Preocupado/a | 4 | 4 | 5 | 4 | 4 | 5 | 4 | 5 | 5 | 4 | 5 | 4 | 5 | 4 | 5 | 4 | 3 | 3 | 5 | 4 | 86 |
| Desesperado/a | 4 | 4 | 5 | 5 | 3 | 5 | 4 | 4 | 5 | 4 | 5 | 3 | 5 | 4 | 5 | 3 | 5 | 3 | 4 | 5 | 85 |
| Desmotivado/a | 4 | 5 | 4 | 4 | 3 | 5 | 3 | 3 | 4 | 4 | 5 | 4 | 5 | 5 | 4 | 5 | 4 | 3 | 5 | 5 | 84 |
| Solo/a | 4 | 4 | 5 | 4 | 4 | 5 | 4 | 4 | 4 | 4 | 5 | 3 | 5 | 4 | 5 | 3 | 4 | 4 | 4 | 5 | 84 |
| Desilusionado/a | 4 | 5 | 3 | 4 | 3 | 5 | 5 | 3 | 4 | 3 | 5 | 4 | 4 | 4 | 5 | 4 | 5 | 4 | 4 | 4 | 82 |
| Incapaz | 4 | 4 | 2 | 4 | 4 | 4 | 4 | 5 | 5 | 3 | 5 | 3 | 5 | 4 | 4 | 5 | 5 | 3 | 4 | 5 | 82 |
| Pena | 4 | 3 | 4 | 3 | 4 | 5 | 3 | 4 | 5 | 5 | 4 | 3 | 5 | 5 | 5 | 4 | 4 | 4 | 3 | 5 | 82 |
| Intranquilo/a | 5 | 4 | 4 | 5 | 4 | 4 | 4 | 3 | 5 | 5 | 3 | 3 | 2 | 4 | 4 | 5 | 4 | 3 | 5 | 5 | 81 |
| Vacío/a | 4 | 5 | 5 | 4 | 3 | 2 | 4 | 4 | 3 | 5 | 5 | 5 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 81 |
| Fatigado/a | 4 | 5 | 5 | 4 | 3 | 5 | 3 | 2 | 5 | 4 | 2 | 4 | 3 | 4 | 5 | 4 | 5 | 4 | 4 | 5 | 80 |
| Sueño | 3 | 4 | 4 | 5 | 5 | 4 | 2 | 4 | 5 | 4 | 5 | 4 | 3 | 5 | 5 | 4 | 2 | 4 | 4 | 4 | 80 |
| Sufrimiento | 3 | 4 | 4 | 5 | 4 | 5 | 4 | 3 | 5 | 2 | 5 | 5 | 4 | 3 | 5 | 4 | 4 | 3 | 3 | 5 | 80 |
| Desánimo | 5 | 5 | 2 | 4 | 4 | 5 | 5 | 3 | 4 | 3 | 3 | 5 | 4 | 5 | 5 | 3 | 4 | 2 | 3 | 5 | 79 |
| Irritable | 3 | 4 | 4 | 4 | 3 | 4 | 5 | 3 | 4 | 5 | 5 | 2 | 4 | 4 | 3 | 5 | 5 | 4 | 4 | 4 | 79 |
| Culpa | 3 | 4 | 5 | 4 | 4 | 3 | 5 | 4 | 4 | 3 | 5 | 3 | 5 | 4 | 5 | 3 | 3 | 2 | 4 | 5 | 78 |
| Culpable | 3 | 4 | 5 | 3 | 4 | 3 | 5 | 4 | 4 | 3 | 5 | 3 | 5 | 4 | 5 | 3 | 3 | 3 | 4 | 5 | 78 |
| Desilusión | 4 | 4 | 2 | 5 | 3 | 5 | 5 | 3 | 3 | 3 | 4 | 4 | 4 | 4 | 5 | 4 | 5 | 4 | 3 | 4 | 78 |
| Fracasado/a | 3 | 5 | 3 | 3 | 3 | 4 | 4 | 4 | 3 | 4 | 5 | 4 | 4 | 4 | 4 | 4 | 4 | 3 | 5 | 5 | 78 |
| Hundido/a | 3 | 4 | 3 | 4 | 4 | 4 | 4 | 4 | 3 | 4 | 5 | 3 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 5 | 78 |
| Incomprendido | 3 | 4 | 4 | 4 | 3 | 5 | 4 | 4 | 5 | 3 | 4 | 3 | 4 | 4 | 4 | 4 | 4 | 3 | 5 | 4 | 78 |

[a] Psychiatrist
[b] Family physician

| Words | P[a]1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 | P9 | P10 | P11 | P12 | P13 | P14 | P15 | F[b]1 | F2 | F3 | F4 | F5 | Score |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Preocupación | 4 | 5 | 5 | 3 | 4 | 4 | 4 | 4 | 5 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 3 | 2 | 3 | 4 | 78 |
| Desganado/a | 2 | 4 | 4 | 4 | 3 | 5 | 4 | 4 | 4 | 3 | 5 | 3 | 3 | 4 | 4 | 3 | 5 | 4 | 5 | 4 | 77 |
| Enfermo/a | 4 | 5 | 5 | 4 | 3 | 4 | 2 | 4 | 5 | 3 | 4 | 4 | 5 | 4 | 4 | 2 | 4 | 2 | 4 | 5 | 77 |
| Fatiga | 4 | 5 | 5 | 4 | 3 | 5 | 3 | 2 | 5 | 2 | 2 | 4 | 4 | 4 | 5 | 4 | 5 | 4 | 2 | 5 | 77 |
| Inútil | 4 | 4 | 3 | 5 | 3 | 4 | 4 | 4 | 3 | 4 | 5 | 3 | 5 | 4 | 5 | 4 | 3 | 2 | 4 | 4 | 77 |
| Malestar | 3 | 5 | 5 | 5 | 2 | 5 | 5 | 5 | 5 | 3 | 2 | 3 | 5 | 4 | 4 | 2 | 4 | 3 | 3 | 4 | 77 |
| Fracaso | 3 | 5 | 3 | 3 | 3 | 4 | 4 | 4 | 4 | 4 | 5 | 4 | 4 | 4 | 4 | 4 | 4 | 2 | 3 | 5 | 76 |
| Infeliz | 4 | 3 | 5 | 5 | 2 | 4 | 5 | 4 | 3 | 3 | 4 | 3 | 2 | 3 | 3 | 5 | 5 | 4 | 5 | 4 | 76 |
| Apagado/a | 4 | 3 | 4 | 2 | 4 | 4 | 3 | 3 | 5 | 5 | 5 | 4 | 4 | 5 | 3 | 3 | 5 | 1 | 4 | 4 | 75 |
| Débil | 3 | 4 | 3 | 3 | 3 | 4 | 5 | 4 | 4 | 5 | 5 | 4 | 2 | 5 | 4 | 3 | 5 | 2 | 3 | 4 | 75 |
| Desinterés | 4 | 5 | 4 | 4 | 3 | 4 | 4 | 2 | 2 | 4 | 4 | 4 | 4 | 4 | 2 | 5 | 3 | 3 | 5 | 5 | 75 |
| Irritado/a | 3 | 4 | 4 | 4 | 3 | 2 | 5 | 4 | 4 | 5 | 4 | 4 | 4 | 4 | 1 | 5 | 5 | 3 | 3 | 4 | 75 |
| Lloroso | 3 | 5 | 4 | 4 | 2 | 5 | 5 | 2 | 4 | 3 | 2 | 4 | 4 | 4 | 4 | 5 | 4 | 3 | 4 | 4 | 75 |
| Muerte | 3 | 4 | 4 | 4 | 4 | 4 | 2 | 4 | 4 | 3 | 5 | 4 | 5 | 5 | 5 | 3 | 3 | 3 | 3 | 3 | 75 |
| Peor | 4 | 5 | 4 | 4 | 3 | 3 | 3 | 4 | 5 | 5 | 5 | 3 | 2 | 5 | 5 | 4 | 5 | 1 | 3 | 2 | 75 |
| Desgraciado/a | 3 | 4 | 3 | 4 | 2 | 5 | 4 | 3 | 5 | 3 | 5 | 3 | 3 | 4 | 4 | 4 | 3 | 3 | 4 | 5 | 74 |
| Distraído/a | 4 | 4 | 4 | 4 | 3 | 5 | 3 | 3 | 4 | 4 | 5 | 3 | 4 | 3 | 4 | 4 | 2 | 4 | 4 | 3 | 74 |
| Estorbo | 3 | 4 | 3 | 4 | 3 | 5 | 2 | 4 | 5 | 4 | 4 | 3 | 4 | 4 | 4 | 3 | 5 | 2 | 4 | 4 | 74 |
| Pesimista | 4 | 3 | 3 | 3 | 3 | 4 | 5 | 3 | 4 | 4 | 5 | 3 | 3 | 4 | 3 | 3 | 5 | 3 | 4 | 5 | 74 |
| Desgracia | 3 | 4 | 4 | 4 | 2 | 5 | 4 | 3 | 4 | 3 | 5 | 3 | 5 | 3 | 4 | 4 | 3 | 2 | 3 | 5 | 73 |

[a] Psychiatrist
[b] Family physician

| Words | P[a]1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 | P9 | P10 | P11 | P12 | P13 | P14 | P15 | F[b]1 | F2 | F3 | F4 | F5 | Score |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Bloqueado/a | 3 | 3 | 4 | 4 | 2 | 5 | 4 | 3 | 3 | 3 | 3 | 4 | 5 | 3 | 3 | 3 | 4 | 4 | 4 | 5 | 72 |
| Inquieto/a | 3 | 4 | 5 | 4 | 3 | 4 | 3 | 3 | 4 | 5 | 5 | 5 | 2 | 4 | 3 | 3 | 5 | 2 | 2 | 3 | 72 |
| Inseguro/a | 3 | 4 | 4 | 4 | 4 | 2 | 5 | 3 | 5 | 4 | 5 | 3 | 2 | 5 | 4 | 3 | 2 | 3 | 4 | 3 | 72 |
| Sensible | 3 | 3 | 4 | 3 | 3 | 4 | 4 | 3 | 3 | 3 | 4 | 4 | 4 | 4 | 5 | 3 | 4 | 4 | 3 | 4 | 72 |
| Soledad | 3 | 4 | 4 | 4 | 4 | 5 | 4 | 3 | 4 | 2 | 5 | 3 | 3 | 4 | 3 | 3 | 4 | 3 | 2 | 5 | 72 |
| Culpabilidad | 3 | 4 | 5 | 4 | 3 | 2 | 4 | 3 | 4 | 3 | 5 | 2 | 5 | 2 | 5 | 3 | 3 | 2 | 4 | 5 | 71 |
| Desconcentrado/a | 3 | 5 | 3 | 3 | 3 | 4 | 4 | 4 | 4 | 1 | 3 | 3 | 5 | 4 | 4 | 4 | 4 | 4 | 2 | 4 | 71 |
| Enfermedad | 4 | 5 | 4 | 5 | 3 | 4 | 2 | 3 | 5 | 3 | 5 | 3 | 3 | 5 | 5 | 2 | 4 | 1 | 1 | 4 | 71 |
| Suicidio | 2 | 4 | 3 | 4 | 3 | 5 | 4 | 3 | 4 | 4 | 5 | 4 | 5 | 5 | 3 | 2 | 3 | 3 | 3 | 2 | 71 |
| Dificultad | 3 | 4 | 5 | 4 | 4 | 4 | 3 | 4 | 5 | 3 | 3 | 3 | 5 | 2 | 4 | 3 | 4 | 1 | 2 | 4 | 70 |
| Flojo/a | 3 | 4 | 5 | 4 | 3 | 5 | 3 | 3 | 3 | 4 | 4 | 4 | 2 | 4 | 5 | 3 | 4 | 3 | 2 | 2 | 70 |
| Lento/a | 4 | 3 | 5 | 4 | 3 | 5 | 5 | 4 | 2 | 4 | 5 | 3 | 3 | 4 | 2 | 2 | 2 | 3 | 3 | 4 | 70 |
| Desinteresado/a | 4 | 5 | 3 | 4 | 3 | 5 | 4 | 2 | 3 | 2 | 3 | 4 | 3 | 4 | 1 | 5 | 3 | 2 | 4 | 5 | 69 |
| Frustrado | 3 | 4 | 3 | 4 | 3 | 5 | 2 | 3 | 4 | 4 | 5 | 3 | 3 | 3 | 2 | 4 | 4 | 2 | 3 | 5 | 69 |
| Indeciso/a | 3 | 5 | 4 | 4 | 3 | 4 | 3 | 4 | 3 | 3 | 5 | 3 | 3 | 4 | 3 | 4 | 2 | 2 | 3 | 4 | 69 |
| Avergonzado/a | 3 | 3 | 4 | 4 | 3 | 5 | 4 | 3 | 4 | 3 | 4 | 3 | 4 | 3 | 4 | 3 | 2 | 2 | 3 | 4 | 68 |
| Decaimiento | 4 | 4 | 2 | 3 | 4 | 2 | 4 | 3 | 3 | 5 | 3 | 3 | 3 | 5 | 2 | 4 | 4 | 2 | 3 | 5 | 68 |
| Decepcionado/a | 3 | 3 | 3 | 3 | 3 | 4 | 2 | 4 | 4 | 3 | 5 | 2 | 5 | 4 | 3 | 4 | 2 | 2 | 4 | 5 | 68 |
| Descentrado/a | 3 | 5 | 3 | 4 | 3 | 4 | 4 | 3 | 3 | 2 | 2 | 3 | 4 | 4 | 4 | 2 | 4 | 3 | 4 | 4 | 68 |
| Disperso/a | 3 | 4 | 5 | 4 | 3 | 4 | 2 | 4 | 3 | 2 | 4 | 2 | 4 | 4 | 4 | 4 | 2 | 3 | 4 | 3 | 68 |

[a] Psychiatrist
[b] Family physician

| Words | P[a]1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 | P9 | P10 | P11 | P12 | P13 | P14 | P15 | F[b]1 | F2 | F3 | F4 | F5 | Score |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Confundido/a | 2 | 4 | 5 | 2 | 3 | 5 | 3 | 3 | 3 | 4 | 5 | 3 | 3 | 4 | 1 | 3 | 2 | 2 | 3 | 4 | 64 |
| Desesperanzado/a | 3 | 5 | 2 | 3 | 3 | 5 | 4 | 1 | 2 | 4 | 2 | 3 | 4 | 4 | 2 | 3 | 5 | 2 | 2 | 5 | 64 |
| Dolorido/a | 3 | 4 | 4 | 4 | 2 | 4 | 2 | 3 | 4 | 2 | 3 | 2 | 3 | 4 | 5 | 3 | 4 | 2 | 2 | 4 | 64 |
| Incapacidad | 4 | 4 | 2 | 3 | 4 | 3 | 4 | 2 | 4 | 3 | 2 | 3 | 2 | 3 | 5 | 4 | 4 | 2 | 2 | 4 | 64 |
| Inestable | 3 | 4 | 4 | 4 | 3 | 2 | 3 | 3 | 4 | 4 | 3 | 2 | 2 | 4 | 3 | 3 | 4 | 2 | 4 | 3 | 64 |
| Mareado | 2 | 4 | 4 | 5 | 2 | 3 | 3 | 3 | 4 | 2 | 4 | 3 | 3 | 4 | 4 | 4 | 3 | 2 | 3 | 2 | 64 |
| Temor | 3 | 4 | 2 | 4 | 3 | 4 | 2 | 2 | 3 | 4 | 3 | 3 | 4 | 4 | 3 | 4 | 4 | 1 | 3 | 4 | 64 |
| Autolesionarse | 3 | 3 | 3 | 3 | 4 | 4 | 5 | 3 | 2 | 4 | 4 | 3 | 5 | 3 | 3 | 2 | 3 | 2 | 2 | 2 | 63 |
| Desmoralizado | 4 | 3 | 1 | 4 | 1 | 2 | 2 | 1 | 3 | 4 | 4 | 1 | 4 | 4 | 4 | 4 | 4 | 3 | 5 | 5 | 63 |
| Inactivo/a | 4 | 4 | 5 | 5 | 2 | 1 | 3 | 2 | 2 | 4 | 4 | 3 | 2 | 4 | 3 | 4 | 4 | 1 | 2 | 4 | 63 |
| Infelicidad | 3 | 4 | 4 | 5 | 2 | 4 | 5 | 2 | 2 | 3 | 4 | 3 | 2 | 2 | 1 | 3 | 4 | 3 | 3 | 4 | 63 |
| Miedoso/a | 3 | 4 | 3 | 4 | 3 | 4 | 3 | 3 | 4 | 3 | 5 | 3 | 3 | 3 | 3 | 2 | 2 | 2 | 3 | 3 | 63 |
| Decepción | 3 | 3 | 3 | 3 | 3 | 2 | 2 | 3 | 4 | 2 | 4 | 2 | 5 | 4 | 3 | 4 | 3 | 1 | 3 | 5 | 62 |
| Olvidadizo/a | 3 | 4 | 3 | 4 | 4 | 4 | 3 | 2 | 4 | 1 | 2 | 2 | 3 | 3 | 2 | 3 | 3 | 4 | 4 | 4 | 62 |
| Aislado/a | 2 | 3 | 3 | 2 | 2 | 4 | 4 | 3 | 3 | 4 | 4 | 2 | 4 | 5 | 2 | 2 | 3 | 3 | 2 | 4 | 61 |
| Desagradable | 1 | 3 | 3 | 4 | 4 | 4 | 3 | 4 | 4 | 2 | 3 | 3 | 4 | 4 | 4 | 2 | 2 | 1 | 3 | 3 | 61 |
| Descuidado/a | 3 | 4 | 3 | 4 | 2 | 3 | 2 | 3 | 3 | 3 | 3 | 3 | 5 | 4 | 2 | 1 | 2 | 2 | 5 | 4 | 61 |
| Alterado/a | 3 | 4 | 5 | 2 | 2 | 1 | 4 | 3 | 3 | 3 | 4 | 3 | 3 | 4 | 2 | 3 | 3 | 2 | 2 | 4 | 60 |
| Derrumbado/a | 2 | 3 | 2 | 4 | 2 | 2 | 3 | 2 | 3 | 2 | 4 | 3 | 3 | 3 | 3 | 3 | 5 | 3 | 3 | 5 | 60 |
| Inferioridad | 2 | 3 | 2 | 4 | 3 | 2 | 5 | 3 | 2 | 3 | 4 | 3 | 5 | 2 | 3 | 3 | 4 | 1 | 4 | 2 | 60 |

[a] Psychiatrist
[b] Family physician

| Words | P[a]1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 | P9 | P10 | P11 | P12 | P13 | P14 | P15 | F[b]1 | F2 | F3 | F4 | F5 | Score |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Desconectado/a | 2 | 5 | 3 | 3 | 2 | 3 | 3 | 2 | 3 | 4 | 3 | 2 | 3 | 3 | 2 | 4 | 4 | 2 | 2 | 4 | 59 |
| Acabado/a | 3 | 2 | 3 | 2 | 1 | 3 | 4 | 3 | 3 | 2 | 4 | 3 | 3 | 3 | 2 | 3 | 4 | 3 | 4 | 3 | 58 |
| Estúpido/a | 2 | 4 | 3 | 4 | 3 | 3 | 3 | 3 | 4 | 4 | 4 | 2 | 3 | 3 | 2 | 3 | 2 | 1 | 2 | 3 | 58 |
| Solitario/a | 2 | 3 | 4 | 3 | 3 | 4 | 4 | 3 | 2 | 2 | 3 | 2 | 2 | 3 | 1 | 3 | 4 | 2 | 4 | 4 | 58 |
| Desastre | 3 | 2 | 3 | 4 | 2 | 2 | 2 | 3 | 3 | 2 | 4 | 2 | 4 | 3 | 3 | 3 | 4 | 2 | 2 | 4 | 57 |
| Indiferente | 3 | 3 | 3 | 3 | 2 | 4 | 3 | 2 | 2 | 2 | 4 | 3 | 3 | 3 | 2 | 4 | 3 | 2 | 3 | 3 | 57 |
| Pasivo/a | 2 | 3 | 4 | 3 | 3 | 4 | 3 | 3 | 2 | 3 | 4 | 1 | 3 | 3 | 2 | 3 | 2 | 1 | 4 | 4 | 57 |
| Suicida | 3 | 3 | 3 | 4 | 2 | 5 | 4 | 2 | 2 | 2 | 4 | 3 | 2 | 3 | 3 | 2 | 3 | 2 | 3 | 2 | 57 |
| Derrotado/a | 3 | 3 | 2 | 3 | 2 | 3 | 2 | 2 | 4 | 2 | 5 | 2 | 2 | 3 | 1 | 2 | 5 | 1 | 4 | 5 | 56 |
| Enlentecido/a | 3 | 4 | 2 | 3 | 2 | 2 | 2 | 3 | 2 | 3 | 2 | 3 | 4 | 3 | 2 | 3 | 4 | 3 | 2 | 4 | 56 |
| Ido/a | 2 | 3 | 3 | 3 | 4 | 1 | 2 | 3 | 3 | 2 | 4 | 2 | 4 | 4 | 1 | 4 | 2 | 3 | 3 | 3 | 56 |
| Insignificante | 2 | 3 | 3 | 4 | 3 | 3 | 4 | 3 | 3 | 3 | 3 | 2 | 2 | 2 | 2 | 2 | 3 | 3 | 4 | 2 | 56 |
| Abatido/a | 2 | 5 | 2 | 2 | 1 | 4 | 3 | 1 | 2 | 3 | 3 | 4 | 3 | 1 | 1 | 4 | 4 | 3 | 4 | 3 | 55 |
| Asustado/a | 3 | 3 | 2 | 2 | 2 | 4 | 3 | 3 | 5 | 3 | 4 | 2 | 3 | 3 | 3 | 2 | 1 | 1 | 3 | 3 | 55 |
| Criticado/a | 2 | 3 | 4 | 3 | 3 | 1 | 3 | 4 | 2 | 2 | 4 | 2 | 4 | 3 | 2 | 2 | 4 | 1 | 3 | 3 | 55 |
| Desconfiado/a | 3 | 3 | 3 | 3 | 3 | 2 | 3 | 3 | 4 | 3 | 5 | 2 | 3 | 3 | 1 | 3 | 3 | 1 | 2 | 2 | 55 |
| Desengaño | 2 | 2 | 2 | 4 | 2 | 3 | 2 | 3 | 3 | 3 | 5 | 1 | 2 | 3 | 3 | 3 | 3 | 2 | 3 | 4 | 55 |
| Desvelado/a | 3 | 3 | 2 | 4 | 4 | 4 | 1 | 2 | 4 | 5 | 1 | 2 | 1 | 5 | 4 | 1 | 2 | 2 | 1 | 4 | 55 |
| Ignorado/a | 2 | 3 | 3 | 3 | 4 | 2 | 2 | 2 | 4 | 2 | 3 | 2 | 3 | 3 | 3 | 3 | 4 | 2 | 3 | 2 | 55 |
| Inapetente | 3 | 3 | 2 | 4 | 2 | 2 | 3 | 4 | 2 | 2 | 2 | 3 | 1 | 2 | 3 | 3 | 5 | 2 | 3 | 4 | 55 |

[a] Psychiatrist
[b] Family physician

| Words | P[a]1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 | P9 | P10 | P11 | P12 | P13 | P14 | P15 | F[b]1 | F2 | F3 | F4 | F5 | Score |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Insensible | 2 | 4 | 3 | 3 | 2 | 2 | 3 | 3 | 3 | 3 | 3 | 2 | 2 | 4 | 1 | 3 | 3 | 2 | 5 | 2 | 55 |
| Pobre | 2 | 3 | 3 | 3 | 3 | 2 | 2 | 2 | 4 | 2 | 5 | 3 | 4 | 4 | 1 | 3 | 3 | 1 | 3 | 2 | 55 |
| Torpe | 2 | 4 | 3 | 3 | 3 | 5 | 2 | 4 | 3 | 4 | 2 | 2 | 2 | 2 | 3 | 2 | 3 | 2 | 2 | 2 | 55 |
| Víctima | 1 | 3 | 4 | 3 | 3 | 4 | 3 | 3 | 2 | 4 | 3 | 3 | 3 | 3 | 1 | 3 | 3 | 1 | 2 | 3 | 55 |
| Apenado/a | 4 | 3 | 3 | 2 | 2 | 3 | 4 | 2 | 2 | 3 | 2 | 3 | 3 | 2 | 1 | 3 | 3 | 2 | 2 | 5 | 54 |
| Ausente | 3 | 3 | 3 | 2 | 3 | 4 | 2 | 3 | 3 | 3 | 2 | 3 | 2 | 2 | 2 | 2 | 1 | 4 | 3 | 4 | 54 |
| Desconsolado/a | 4 | 3 | 2 | 4 | 2 | 1 | 1 | 2 | 4 | 3 | 4 | 1 | 2 | 2 | 3 | 2 | 5 | 2 | 4 | 3 | 54 |
| Incompetente | 1 | 4 | 2 | 4 | 3 | 2 | 3 | 2 | 2 | 3 | 2 | 2 | 4 | 1 | 4 | 4 | 3 | 2 | 3 | 3 | 54 |
| Tímido/a | 2 | 3 | 4 | 3 | 3 | 1 | 3 | 3 | 2 | 3 | 4 | 3 | 2 | 4 | 1 | 3 | 3 | 1 | 4 | 2 | 54 |
| Vergonzoso | 3 | 3 | 4 | 2 | 3 | 3 | 2 | 3 | 3 | 4 | 2 | 2 | 1 | 3 | 3 | 3 | 4 | 2 | 2 | 2 | 54 |
| Abandonado/a | 1 | 3 | 3 | 4 | 2 | 2 | 2 | 2 | 3 | 3 | 2 | 2 | 4 | 3 | 3 | 3 | 2 | 2 | 4 | 3 | 53 |
| Autorreproches | 4 | 4 | 2 | 3 | 1 | 4 | 4 | 2 | 3 | 2 | 3 | 2 | 1 | 2 | 1 | 3 | 3 | 2 | 3 | 4 | 53 |
| Inestabilidad | 2 | 3 | 4 | 4 | 3 | 2 | 3 | 2 | 2 | 3 | 2 | 1 | 2 | 4 | 2 | 2 | 4 | 2 | 3 | 3 | 53 |
| Paralizado/a | 2 | 3 | 2 | 4 | 3 | 2 | 3 | 3 | 2 | 3 | 3 | 3 | 2 | 3 | 2 | 3 | 2 | 2 | 2 | 4 | 53 |
| Perdedor/a | 2 | 3 | 2 | 2 | 3 | 2 | 2 | 3 | 2 | 2 | 4 | 2 | 2 | 3 | 2 | 4 | 3 | 1 | 4 | 5 | 53 |
| Castigo | 3 | 2 | 3 | 3 | 3 | 2 | 2 | 3 | 3 | 2 | 5 | 2 | 1 | 4 | 3 | 1 | 3 | 1 | 2 | 4 | 52 |
| Cobarde | 2 | 3 | 3 | 3 | 2 | 2 | 2 | 4 | 3 | 3 | 5 | 2 | 2 | 4 | 2 | 1 | 2 | 3 | 2 | 2 | 52 |
| Desconcertado/a | 2 | 5 | 3 | 2 | 2 | 2 | 4 | 1 | 3 | 3 | 3 | 1 | 3 | 3 | 1 | 4 | 3 | 1 | 2 | 4 | 52 |
| Desorientado/a | 2 | 4 | 3 | 3 | 1 | 1 | 2 | 3 | 3 | 3 | 4 | 2 | 3 | 3 | 2 | 3 | 2 | 2 | 2 | 4 | 52 |
| Distante | 2 | 3 | 3 | 3 | 3 | 4 | 2 | 3 | 2 | 3 | 4 | 2 | 3 | 2 | 1 | 3 | 2 | 2 | 3 | 2 | 52 |

[a] Psychiatrist
[b] Family physician

| Words | P[a]1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 | P9 | P10 | P11 | P12 | P13 | P14 | P15 | F[b]1 | F2 | F3 | F4 | F5 | Score |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Error | 2 | 3 | 3 | 3 | 3 | 1 | 3 | 4 | 5 | 2 | 5 | 2 | 4 | 2 | 1 | 2 | 2 | 2 | 1 | 2 | 52 |
| Indefenso | 2 | 3 | 2 | 4 | 2 | 1 | 3 | 2 | 3 | 3 | 2 | 3 | 2 | 3 | 3 | 3 | 3 | 2 | 4 | 2 | 52 |
| Vencido | 3 | 3 | 1 | 3 | 3 | 2 | 2 | 3 | 3 | 2 | 3 | 2 | 1 | 2 | 1 | 3 | 4 | 2 | 4 | 5 | 52 |
| Amargura | 3 | 3 | 1 | 3 | 3 | 3 | 4 | 1 | 2 | 3 | 3 | 1 | 2 | 3 | 1 | 3 | 3 | 3 | 1 | 4 | 50 |
| Apatía | 4 | 2 | 3 | 1 | 1 | 2 | 2 | 1 | 1 | 4 | 2 | 3 | 3 | 4 | 3 | 2 | 4 | 1 | 3 | 4 | 50 |
| Defraudado | 2 | 2 | 2 | 2 | 2 | 4 | 2 | 2 | 3 | 2 | 4 | 1 | 4 | 3 | 2 | 2 | 3 | 1 | 3 | 4 | 50 |
| Destruido/a | 2 | 2 | 3 | 4 | 2 | 3 | 2 | 2 | 2 | 3 | 4 | 2 | 2 | 2 | 2 | 2 | 4 | 1 | 3 | 3 | 50 |
| Introvertido/a | 2 | 3 | 4 | 4 | 3 | 1 | 2 | 2 | 4 | 3 | 3 | 2 | 2 | 2 | 1 | 4 | 3 | 1 | 2 | 2 | 50 |
| Terror | 2 | 3 | 2 | 2 | 2 | 2 | 3 | 2 | 2 | 3 | 5 | 3 | 4 | 2 | 1 | 3 | 3 | 1 | 3 | 2 | 50 |
| Desconfianza | 3 | 3 | 3 | 3 | 2 | 2 | 3 | 2 | 3 | 3 | 3 | 2 | 2 | 3 | 1 | 3 | 3 | 1 | 2 | 2 | 49 |
| Abatimiento | 2 | 4 | 4 | 2 | 1 | 4 | 3 | 1 | 2 | 1 | 2 | 3 | 2 | 2 | 1 | 4 | 3 | 2 | 2 | 3 | 48 |
| Desafortunado/a | 2 | 3 | 2 | 3 | 2 | 3 | 2 | 1 | 3 | 2 | 2 | 2 | 2 | 2 | 2 | 3 | 4 | 1 | 2 | 5 | 48 |
| Inservible | 1 | 4 | 1 | 3 | 3 | 2 | 2 | 4 | 3 | 3 | 2 | 1 | 1 | 1 | 1 | 3 | 4 | 3 | 3 | 3 | 48 |
| Ridículo | 2 | 2 | 3 | 3 | 3 | 2 | 3 | 2 | 2 | 2 | 4 | 2 | 4 | 2 | 2 | 2 | 3 | 1 | 2 | 2 | 48 |
| Tonto | 1 | 3 | 2 | 3 | 3 | 2 | 2 | 3 | 3 | 2 | 5 | 3 | 3 | 4 | 3 | 2 | 1 | 1 | 1 | 1 | 48 |
| Desamparado/a | 2 | 4 | 1 | 4 | 2 | 1 | 2 | 1 | 2 | 2 | 2 | 1 | 4 | 3 | 2 | 2 | 3 | 1 | 4 | 4 | 47 |
| Herido/a | 2 | 2 | 3 | 3 | 2 | 1 | 3 | 3 | 3 | 3 | 2 | 1 | 3 | 3 | 1 | 2 | 3 | 1 | 2 | 4 | 47 |
| Melancólico/ca | 3 | 2 | 3 | 3 | 2 | 2 | 1 | 2 | 1 | 3 | 1 | 3 | 3 | 3 | 1 | 3 | 3 | 1 | 2 | 5 | 47 |
| Ofendido/a | 2 | 2 | 2 | 3 | 3 | 2 | 2 | 2 | 2 | 3 | 4 | 2 | 2 | 2 | 2 | 3 | 3 | 1 | 3 | 2 | 47 |
| Retraído/a | 3 | 2 | 3 | 2 | 2 | 1 | 2 | 1 | 2 | 4 | 5 | 2 | 2 | 3 | 1 | 3 | 3 | 1 | 2 | 3 | 47 |

[a] Psychiatrist
[b] Family physician

| Words | P[a]1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 | P9 | P10 | P11 | P12 | P13 | P14 | P15 | F[b]1 | F2 | F3 | F4 | F5 | Score |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Desatendido | 3 | 2 | 1 | 2 | 3 | 4 | 2 | 2 | 3 | 2 | 2 | 2 | 2 | 1 | 3 | 2 | 3 | 1 | 2 | 4 | 46 |
| Desencantado/a | 2 | 2 | 2 | 4 | 2 | 1 | 1 | 3 | 3 | 2 | 2 | 1 | 1 | 3 | 3 | 3 | 4 | 2 | 3 | 2 | 46 |
| Inhibido/a | 3 | 3 | 2 | 3 | 2 | 1 | 3 | 2 | 1 | 2 | 3 | 2 | 2 | 2 | 1 | 4 | 5 | 1 | 2 | 2 | 46 |
| Miserable | 3 | 2 | 3 | 3 | 3 | 3 | 2 | 1 | 2 | 3 | 3 | 2 | 2 | 2 | 1 | 2 | 2 | 1 | 3 | 3 | 46 |
| Catástrofe | 3 | 3 | 3 | 3 | 2 | 1 | 2 | 2 | 3 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 3 | 1 | 2 | 4 | 45 |
| Desprecio | 1 | 2 | 3 | 3 | 3 | 2 | 2 | 3 | 2 | 3 | 4 | 2 | 2 | 2 | 2 | 1 | 3 | 1 | 2 | 2 | 45 |
| Humillación | 2 | 2 | 3 | 3 | 3 | 2 | 3 | 2 | 2 | 2 | 4 | 1 | 3 | 2 | 2 | 1 | 2 | 1 | 2 | 3 | 45 |
| Agitación | 3 | 3 | 1 | 2 | 1 | 2 | 3 | 3 | 2 | 4 | 2 | 2 | 1 | 3 | 1 | 2 | 2 | 2 | 4 | 1 | 44 |
| Desconsuelo | 4 | 3 | 1 | 4 | 1 | 1 | 1 | 1 | 2 | 1 | 2 | 1 | 3 | 2 | 2 | 2 | 5 | 1 | 3 | 4 | 44 |
| Desmoralización | 3 | 3 | 1 | 3 | 1 | 2 | 2 | 1 | 1 | 4 | 2 | 1 | 3 | 1 | 1 | 4 | 4 | 1 | 2 | 4 | 44 |
| Agonía | 2 | 3 | 4 | 3 | 1 | 1 | 2 | 1 | 2 | 2 | 2 | 2 | 1 | 2 | 3 | 2 | 2 | 2 | 2 | 4 | 43 |
| Aturdido/a | 2 | 3 | 2 | 2 | 2 | 1 | 3 | 2 | 2 | 3 | 3 | 2 | 2 | 4 | 1 | 3 | 1 | 1 | 2 | 2 | 43 |
| Desamparo | 2 | 4 | 1 | 4 | 2 | 1 | 2 | 1 | 2 | 2 | 1 | 1 | 2 | 3 | 2 | 2 | 3 | 1 | 3 | 4 | 43 |
| Hipocondría | 3 | 3 | 3 | 3 | 1 | 2 | 2 | 2 | 1 | 2 | 5 | 2 | 1 | 3 | 2 | 1 | 2 | 1 | 2 | 2 | 43 |
| Melancolía | 3 | 2 | 2 | 3 | 2 | 1 | 1 | 2 | 1 | 2 | 1 | 2 | 3 | 3 | 1 | 2 | 4 | 1 | 2 | 5 | 43 |
| Patético/a | 1 | 2 | 3 | 3 | 3 | 3 | 2 | 2 | 2 | 2 | 2 | 2 | 4 | 2 | 1 | 1 | 1 | 1 | 3 | 3 | 43 |
| Alucinaciones | 1 | 2 | 3 | 2 | 1 | 4 | 2 | 3 | 1 | 2 | 4 | 2 | 1 | 4 | 2 | 2 | 1 | 1 | 2 | 2 | 42 |
| Derrota | 2 | 2 | 1 | 2 | 1 | 2 | 2 | 1 | 3 | 2 | 2 | 2 | 2 | 1 | 1 | 2 | 5 | 1 | 3 | 5 | 42 |
| Desolado/a | 3 | 2 | 1 | 3 | 2 | 2 | 1 | 1 | 3 | 3 | 3 | 1 | 2 | 2 | 2 | 2 | 2 | 1 | 2 | 4 | 42 |
| Despreciable | 1 | 2 | 2 | 3 | 2 | 2 | 2 | 2 | 4 | 3 | 2 | 2 | 3 | 2 | 2 | 1 | 3 | 1 | 2 | 1 | 42 |

[a] Psychiatrist
[b] Family physician

| Words | P[a]1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 | P9 | P10 | P11 | P12 | P13 | P14 | P15 | F[b]1 | F2 | F3 | F4 | F5 | Score |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Monótono/a | 1 | 2 | 2 | 3 | 3 | 4 | 2 | 3 | 3 | 1 | 3 | 1 | 2 | 1 | 1 | 1 | 2 | 3 | 3 | 1 | 42 |
| Tormento | 2 | 2 | 2 | 2 | 3 | 1 | 2 | 1 | 3 | 2 | 2 | 2 | 3 | 2 | 1 | 2 | 4 | 1 | 1 | 4 | 42 |
| Conformista | 2 | 2 | 2 | 2 | 2 | 1 | 2 | 3 | 2 | 2 | 5 | 1 | 2 | 2 | 1 | 2 | 2 | 1 | 3 | 2 | 41 |
| Desdichado/a | 4 | 2 | 1 | 3 | 2 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 2 | 1 | 2 | 4 | 2 | 4 | 4 | 41 |
| Disminuido/a | 2 | 2 | 3 | 2 | 1 | 4 | 2 | 1 | 4 | 3 | 1 | 2 | 2 | 1 | 1 | 2 | 3 | 1 | 2 | 2 | 41 |
| Enfermizo/a | 2 | 3 | 1 | 3 | 2 | 1 | 2 | 1 | 2 | 3 | 1 | 1 | 1 | 3 | 1 | 1 | 4 | 1 | 4 | 4 | 41 |
| Agarrotamiento | 1 | 3 | 2 | 1 | 1 | 3 | 2 | 1 | 2 | 3 | 2 | 4 | 1 | 2 | 2 | 2 | 1 | 2 | 3 | 2 | 40 |
| Delirio | 3 | 1 | 3 | 3 | 3 | 1 | 3 | 2 | 3 | 3 | 2 | 2 | 2 | 2 | 1 | 1 | 2 | 1 | 1 | 1 | 40 |
| Acomplejado/a | 1 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 3 | 2 | 3 | 3 | 1 | 3 | 2 | 2 | 1 | 2 | 2 | 3 | 39 |
| Derrotista | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 2 | 4 | 1 | 3 | 5 | 39 |
| Feo | 1 | 2 | 1 | 3 | 3 | 2 | 1 | 2 | 3 | 4 | 4 | 2 | 2 | 2 | 1 | 2 | 1 | 1 | 1 | 1 | 39 |
| Ignorante | 2 | 2 | 2 | 3 | 3 | 1 | 2 | 2 | 1 | 2 | 3 | 1 | 1 | 4 | 1 | 2 | 2 | 1 | 2 | 2 | 39 |
| Indigno/a | 2 | 2 | 2 | 2 | 2 | 1 | 3 | 1 | 1 | 3 | 2 | 1 | 3 | 1 | 2 | 3 | 3 | 1 | 2 | 2 | 39 |
| Mediocre | 2 | 3 | 1 | 3 | 3 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 3 | 1 | 4 | 2 | 39 |
| Pesadumbre | 3 | 2 | 1 | 2 | 2 | 2 | 2 | 1 | 2 | 2 | 1 | 1 | 1 | 3 | 1 | 2 | 3 | 1 | 2 | 5 | 39 |
| Afligido/a | 2 | 2 | 1 | 1 | 2 | 1 | 3 | 1 | 1 | 3 | 2 | 1 | 2 | 2 | 1 | 3 | 2 | 1 | 3 | 4 | 38 |
| Autodesprecio | 3 | 2 | 2 | 2 | 1 | 1 | 3 | 1 | 2 | 2 | 3 | 2 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 3 | 38 |
| Desalentado/a | 3 | 2 | 1 | 2 | 2 | 1 | 1 | 1 | 2 | 2 | 2 | 1 | 1 | 2 | 1 | 3 | 3 | 1 | 4 | 3 | 38 |
| Delirante | 3 | 1 | 3 | 2 | 3 | 1 | 3 | 3 | 2 | 3 | 2 | 2 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 37 |
| Desfallecido | 1 | 2 | 1 | 3 | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 1 | 3 | 1 | 3 | 4 | 1 | 2 | 4 | 37 |

[a] Psychiatrist
[b] Family physician

| Words | P[a]1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 | P9 | P10 | P11 | P12 | P13 | P14 | P15 | F[b]1 | F2 | F3 | F4 | F5 | Score |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Entumecido/a | 2 | 4 | 1 | 3 | 2 | 2 | 2 | 1 | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 1 | 4 | 1 | 2 | 3 | 37 |
| Rumiaciones | 3 | 2 | 2 | 3 | 2 | 1 | 2 | 2 | 1 | 1 | 1 | 2 | 1 | 3 | 2 | 1 | 4 | 1 | 1 | 2 | 37 |
| Simple | 2 | 2 | 2 | 1 | 2 | 1 | 2 | 2 | 3 | 1 | 3 | 3 | 1 | 1 | 1 | 2 | 3 | 1 | 2 | 2 | 37 |
| Anhedonia | 4 | 2 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 5 | 1 | 3 | 4 | 36 |
| Hastío | 2 | 2 | 1 | 4 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 1 | 3 | 1 | 1 | 1 | 1 | 3 | 4 | 36 |
| Inepto/a | 1 | 2 | 2 | 3 | 2 | 1 | 3 | 1 | 2 | 2 | 2 | 1 | 3 | 2 | 1 | 3 | 2 | 1 | 1 | 1 | 36 |
| Turbado | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 3 | 1 | 2 | 1 | 2 | 1 | 2 | 2 | 1 | 3 | 4 | 36 |
| Apesadumbrado/a | 4 | 3 | 1 | 2 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 3 | 2 | 1 | 1 | 3 | 35 |
| Condenado/a | 2 | 3 | 2 | 3 | 1 | 1 | 1 | 1 | 2 | 1 | 3 | 1 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 4 | 35 |
| Desdicha | 3 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 2 | 1 | 2 | 4 | 1 | 3 | 4 | 35 |
| Desolación | 3 | 2 | 1 | 3 | 1 | 2 | 2 | 1 | 2 | 2 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 1 | 3 | 35 |
| Imperfecto/a | 1 | 2 | 2 | 2 | 2 | 1 | 3 | 3 | 2 | 2 | 2 | 1 | 2 | 2 | 1 | 1 | 3 | 1 | 1 | 1 | 35 |
| Inadecuado/a | 3 | 3 | 3 | 2 | 2 | 1 | 3 | 1 | 1 | 2 | 3 | 2 | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 1 | 35 |
| Pesaroso/a | 1 | 2 | 1 | 1 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 1 | 1 | 2 | 1 | 3 | 3 | 1 | 3 | 5 | 35 |
| Trágico | 2 | 2 | 1 | 2 | 3 | 1 | 2 | 1 | 2 | 2 | 2 | 1 | 2 | 2 | 1 | 2 | 2 | 1 | 2 | 2 | 35 |
| Acongojado/a | 2 | 3 | 1 | 1 | 1 | 2 | 2 | 1 | 2 | 2 | 1 | 1 | 1 | 3 | 1 | 2 | 3 | 1 | 2 | 2 | 34 |
| Alicaído/a | 3 | 1 | 1 | 2 | 1 | 1 | 3 | 1 | 2 | 1 | 2 | 1 | 1 | 3 | 1 | 2 | 1 | 1 | 3 | 3 | 34 |
| Descorazonado/a | 3 | 2 | 1 | 3 | 2 | 1 | 1 | 1 | 2 | 2 | 2 | 1 | 1 | 2 | 1 | 1 | 3 | 1 | 2 | 2 | 34 |
| Desnutrido/a | 2 | 3 | 2 | 2 | 1 | 1 | 2 | 3 | 2 | 1 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 2 | 2 | 34 |
| Despecho | 1 | 2 | 2 | 3 | 2 | 2 | 1 | 2 | 3 | 1 | 3 | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 1 | 2 | 34 |

[a] Psychiatrist
[b] Family physician

| Words | P[a]1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 | P9 | P10 | P11 | P12 | P13 | P14 | P15 | F[b]1 | F2 | F3 | F4 | F5 | Score |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Consternado/a | 3 | 3 | 1 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 1 | 3 | 3 | 33 |
| Desapego | 3 | 1 | 1 | 2 | 1 | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 2 | 1 | 2 | 4 | 33 |
| Apetencia | 2 | 2 | 1 | 2 | 1 | 1 | 1 | 2 | 1 | 2 | 1 | 1 | 1 | 1 | 3 | 2 | 4 | 1 | 1 | 2 | 32 |
| Deficiente | 1 | 2 | 2 | 2 | 1 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 4 | 1 | 1 | 1 | 2 | 1 | 1 | 3 | 32 |
| Pecado | 1 | 2 | 1 | 2 | 3 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 2 | 2 | 1 | 2 | 2 | 1 | 1 | 1 | 32 |
| Tedio | 1 | 2 | 1 | 2 | 3 | 1 | 2 | 1 | 1 | 2 | 1 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 2 | 4 | 31 |
| Autocrítico | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 3 | 1 | 3 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 2 | 3 | 30 |
| Ingrato | 1 | 2 | 1 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 3 | 1 | 2 | 2 | 30 |
| Mísero/a | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 1 | 2 | 3 | 30 |
| Desatado | 1 | 2 | 1 | 2 | 1 | 1 | 1 | 1 | 2 | 3 | 2 | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 1 | 3 | 29 |
| Languidez | 2 | 1 | 1 | 3 | 2 | 1 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 4 | 1 | 1 | 1 | 29 |
| Mutilado | 1 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 1 | 1 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 1 | 1 | 3 | 28 |
| Apetente | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 4 | 1 | 1 | 1 | 27 |
| Sombrío/a | 1 | 2 | 1 | 1 | 2 | 1 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 27 |
| Achacoso | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 4 | 1 | 1 | 2 | 2 | 26 |
| Desdeñado/a | 2 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 3 | 1 | 1 | 2 | 26 |
| Lasitud | 1 | 1 | 1 | 3 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 22 |

[a] Psychiatrist
[b] Family physician

# 8. BIBLIOGRAPHY

Adrover, C., Bodnar, T., Huang, Z., Telenti, A., & Salathé, M. (2015). Identifying adverse effects of HIV drug treatment and associated sentiments using Twitter. *JMIR Public Health and Surveillance*, *1*(2), e7.

Aichner, T., & Jacob, F. (2015). Measuring the degree of corporate social media use. *International Journal of Market Research*, *57*(2), 257-276.

Alvarez-Mon, M. A., del Barco, A. A., Lahera, G., Quintero, J., Ferre, F., Pereira-Sanchez, V., Ortuño, F., & Alvarez-Mon, M. (2018). Increasing interest of mass communication media and the general public in the distribution of tweets about mental disorders: observational study. *Journal of Medical Internet Research, 20*(5), e205.

Al-Mosaiwi, M., & Johnstone, T. (2018). In an absolute state: elevated use of absolutist words is a marker specific to anxiety, depression, and suicidal ideation. *Clinical Psychological Science: a Journal of the Association for Psychological Science*, *6*(4), 529–542.

Alvaro, N., Conway, M., Doan, S., Lofi, C., Overington, J., & Collier, N. (2015). Crowdsourcing Twitter annotations to identify first-hand experiences of prescription drug use. *Journal of biomedical informatics*, *58*, 280-287.

Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E. D., Gutierrez, J. B., & Kochut, K. (2017). A brief survey of text mining: Classification, clustering and extraction techniques. *arXiv preprint arXiv:1707.02919.*

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Association; 2013.

Anghelescu, I. G., Kohnen, R., Szegedi, A., Klement, S., & Kieser, M. (2006). Comparison of Hypericum extract WS 5570 and paroxetine in ongoing treatment after recovery from an episode of moderate to severe depression: results from a randomized multicenter study. *Pharmacopsychiatry, 39*(6), 213-219.

Arango, C., Díaz-Caneja, C. M., McGorry, P.D., Rapoport, J., Sommer, I.E., Vorstman, J.A., McDaid, D., Marín, O., Serrano-Drozdowskyj, E., Freedman, R., & Carpenter, W. Preventive strategies for mental health. (2018). *The Lancet Psychiatry*, *5*(7), 591-604.

Arseniev-Koehler, A., Lee, H., McCormick, T., & Moreno, M. A. (2016). #Proana: Pro-eating disorder socialization on Twitter. *The Journal of Adolescent health: official* *publication of the Society for Adolescent Medicine, 58*(6), 659-664.

Audeh, B., Calvier, F. E., Bellet, F., Beyens, M. N., Pariente, A., Lillo-Le Louet, A., & Bousquet, C. (2020). Pharmacology and social media: potentials and biases of web forums for drug mention analysis-case study of France. *Health Informatics Journal*, *26*(2), 1253-1272.

Australian Medical Association Council. (2020). *A guide to social media & guide to social media & medical professionalism: the tips and traps every doctor and medical student should know*. https://ama.com.au/sites/default/files/documents/2020%20AMA%20Social%20Media%20Guide%20FINAL_0.pdf

Baingana, F., Al'Absi, M., Becker, A. E., & Pringle, B. (2015). Global research challenges and opportunities for mental health and substance-use disorders. *Nature*, *527*(7578), S172-S177.

Beck, A. T. (1967). *Depression: Clinical, experimental and theoretical aspects*. Harper and Row.

Bellazzi R. (2014). Big data and biomedical informatics: a challenging opportunity. *Yearbook of Medical Informatics*, *9*(1), 8-13.

Ben-Zeev, D., Scherer, E. A., Wang, R., Xie, H., & Campbell, A. T. (2015). Next-generation psychiatric assessment: Using smartphone sensors to monitor behavior and mental health. *Psychiatric Rehabilitation Journal*, *38*(3), 218.

Bernard, J. D., Baddeley, J.L., Rodriguez, B.F., & Burke, P. A. (2016). Depression, language, and affect: an examination of the influence of baseline depression and affect induction on language. *Journal of Language and Social Psychology*, *35*(3), 317-326.

Berry, N., Lobban, F., Belousov, M., Emsley, R., Nenadic, G., & Bucci, S. (2017). # WhyWeTweetMH: understanding why people use Twitter to discuss mental health problems. *Journal of Medical Internet Research*, *19*(4), e107.

Berthon, P. R., Pitt, L. F., Plangger, K., & Shapiro, D. (2012). Marketing meets Web 2.0, social media, and creative consumers: implications for international marketing strategy. *Business Horizons*, *55*(3), 261-271.

Bickman, L. (1996). A continuum of care: More is not always better. *American Psychologist*, *51*(7), 689.

Bidargaddi, N., Musiat, P., Makinen, V. P., Ermes, M., Schrader, G., & Licinio, J. (2017). Digital footprints: facilitating large-scale environmental psychiatric research in naturalistic settings through data from everyday technologies. *Molecular Psychiatry*, *22*(2), 164-169.

Birnbaum, M. L., Ernala, S. K., Rizvi, A. F., De Choudhury, M., & Kane, J. M. (2017). A collaborative approach to identifying social media markers of schizophrenia by employing machine learning and clinical appraisals. *Journal of Medical Internet Research, 19*(8), e289.

Bloom, D.E., Cafiero, E.T., Jané-Llopis, E., Abrahams-Gessel, S., Bloom, L.R., Fathima, S., Feigl, A.B., Gaziano, T., Mowafi, M., Pandya, A., Prettner, K., Rosenberg, L., Seligman, B., Stein, A.Z., & Weinstein, C. (2011). *The global economic burden of non-communicable diseases*. Geneva: World Economic Forum.

Boyd, D. M., & Ellison, N. B. (2007). Social network sites: definition, history, and scholarship. *Journal of Computer-mediated Communication*, *13*(1), 210-230.

British Medical Association. (2018). *Social media, ethics and professionalism.BMA guidance*. https://www.bma.org.uk/media/1851/bma-ethics-guidance-on-social-media-2018.pdf

Brown A. D. (2010). Social media: a new frontier in reflective practice. *Medical Education*, *44*(8), 744-745. https://doi.org/10.1111/j.1365-2923.2010.03729.x

Bucci, W., & Freedman, N. (1981). The language of depression. *Bulletin of the Menninger Clinic*, *45*(4), 334.

Cavazos-Rehg, P. A., Krauss, M. J., Sowles, S., Connolly, S., Rosas, C., Bharadwaj, M., & Bierut, L. J. (2016). A content analysis of depression-related tweets. *Computers in Human Behavior*, *54*, 351-357.

Carbonell, P., Mayer, M. A., & Bravo, À. (2015). Exploring brand-name drug mentions on Twitter for pharmacovigilance. *Studies in health technology and informatics, 210*, 55-59.

Cassano, P., & Fava, M. (2002). Depression and public health: an overview. *Journal of Psychosomatic Research*, *53*(4), 849-857.

Castells, M. (1999). The rise of the network society, the information of the age: Economy, society and culture (Vol. 1). Blackwell Publishers.

Charlson, F. J., Baxter, A. J., Dua, T., Degenhardt, L., Whiteford, H. A., & Vos, T. (2015). Excess mortality from mental, neurological and substance use disorders in the Global Burden of Disease Study 2010. *Epidemiology and Psychiatric Sciences*, *24*(2), 121-140.

Chen, E., Lerman, K., & Ferrara, E. (2020). Tracking social media discourse about the COVID-19 pandemic: development of a public coronavirus Twitter data set. *JMIR Public Health and Surveillance*, *6*(2), e19273.

Chiu, C. Y., Lane, H. Y., Koh, J. L., & Chen, A. L. (2020). Multimodal depression detection on instagram considering time interval of posts. *Journal of Intelligent Information Systems*, 1-23.

Chung, C., & Pennebaker, J. W. (2007). The psychological functions of function words. In: K. Fiedler (Ed.). *Social Communication: Frontiers of Social Psychology* (1st ed., 343-359). Psychology Press.

Clark, L. A., Cuthbert, B., Lewis-Fernández, R., Narrow, W. E., & Reed, G. M. (2017). Three approaches to understanding and classifying mental disorder: ICD-11, DSM-5, and the National Institute of Mental Health's Research Domain Criteria (RDoC). *Psychological Science in the Public Interest*, *18*(2), 72-145.

Clement, J. (2019, August 14). *Number of monthly active Twitter users worldwide from 1st quarter 2010 to 2nd quarter 2019 (in millions)*. Statista. Retrieved July 7, 2020, from https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/

Clement, J. (2019, July 24). *Leading countries based on number of Twitter users as of April 2020 (in millions)*. Statista. Retrieved September 2, 2020, from: https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/

Clement, J. (2020, July 25). *Number of global social network users 2017-2025 (in billions)*. Statista. Retrieved October 29, 2020, from: https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/

Chary, M., Genes, N., McKenzie, A., & Manini, A. F. (2013). Leveraging social networks for toxicovigilance. *Journal of Medical Toxicology, 9*(2), 184-191.

Constantinides, E., & Fountain, S. (2008). Web 2.0: conceptual foundations and marketing issues. *Journal of Direct, Data and Digital Marketing Practice*, *9*(3), 231-244.

Conway, M., & O'Connor, D. (2016). Social media, big data, and mental health: current advances and ethical implications. *Current Opinion in Psychology*, *9*, 77-82.

Coppersmith, G., Dredze, M., & Harman, C. (2014). Quantifying mental health signals in Twitter. *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality* (pp. 51-60). ACL.

Coppersmith, G., Harman, C., & Dredze, M. (2014). Measuring post-traumatic stress disorder in Twitter. *Proceedings of the 8th International Conference on Weblogs and Social Media* (pp. 579-582). AAAI.

Coppersmith, G., Dredze, M., Harman, C., & Hollingshead, K. (2015). From ADHD to SAD: analyzing the language of mental health on twitter through self-reported diagnoses. *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality* (pp. 1-10). ACL.

Crannell, W. C., Clark, E., Jones, C., James, T. A., & Moore, J. (2016). A pattern-matched Twitter analysis of US cancer-patient sentiments. *Journal of Surgical Research*, *206*(2), 536-542.

Cruz, F.L., Troyano, J.A., Pontes, B., & Ortega, F.J. (2014). Building layered, multilingual sentiment lexicons at synset and lemma levels. *Expert Systems With Applications, 41*(13), 5984-5994.

De Choudhury, M., Gamon, M., Count, S., & Horvitz, E. Predicting depression via social media. (2013). *Proceedings of the Seventh International Conference on Weblogs and Social Media* (pp. 128-138). AAAI Press.

De Choudhury, M., Counts, S., & Horvitz, E. (2013). Social Media as a measurement tool of depression in populations. *Proceedings of the 5th Annual ACM Web Science Conference* (pp. 47-56). ACM.

De Choudhury, M., Counts, S., Horvitz, E. J., & Hoff, A. (2014). Characterizing and predicting postpartum depression from shared facebook data. *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing* (pp. 626-638).

De Choudhury, M., & De, S. (2014). Mental health discourse on Reddit: self-disclosure, social Support, and anonymity. *Proceedings of the 8th International Conference on Weblogs and Social Media* (pp. 71-80). AAAI.

De Choudhury, M., Sharma, S. S., Logar, T., Eekhout,W., & Nielsen, R.C. (2017). Gender and cross-cultural differences in social media disclosures of mental illness. *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (pp. 353-369). ACM.

De Choudhury, M., Kiciman, E., Dredze, M., Coppersmith, G., & Kumar, M. (2016). Discovering Shifts to Suicidal Ideation from Mental Health Content in Social Media. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2098-2110). ACM.

De las Cuevas, C., Peñate, W., & Sanz, E. J. (2014). Risk factors for non-adherence to antidepressant treatment in patients with mood disorders. *European Journal of Clinical Pharmacology, 70*(1), 89-98.

Duman, R. S., & Monteggia, L. M. (2006). A neurotrophic model for stress-related mood disorders. *Biological Psychiatry*, *59*(12), 1116-1127.

Durkheim, E. (1951). *Suicide, a study in sociology*. (J. Spaulding & G. Simpson, Trans.). Free Press.

Eichstaedt, J. C., Schwartz, H. A., Kern, M. L., Park, G., Labarthe, D. R., Merchant, R. M., Jha, S., Agrawal, M., Dzjurzynski, L. A., Sap, M., Weeg, C., Larson, E. E., Ungar, L. H., & Seligman, M. E. (2015). Psychological language on Twitter predicts county-level heart disease mortality. *Psychological science, 26*(2), 159-169.

Ekman, P., Friesen, W. V., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., Krause, R., LeCompte, W. A., Pitcairn, T., & Ricci-Bitti, P. E. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of Personality and Social Psychology*, *53*(4), 712-717.

Ennis, L., Rose, D., Denis, M., Pandit, N., & Wykes, T. (2012). Can't surf, won't surf: the digital divide in mental health. *Journal of Mental Health*, *21*(4), 395-403.

Epp, J. (1988). Mental health for Canadians: Striking a balance. Ottawa: Minister of Supplies and Services.

Eysenbach, G. (2009). Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. *Journal of Medical Internet Research*, *11*(1), e11.

Ferrari, A. J., Charlson, F. J., Norman, R. E., Patten, S. B., Freedman, G., Murray, C. J., Vos, T. & Whiteford, H. A.

(2013). Burden of depressive disorders by country, sex, age, and year: findings from the global burden of disease study 2010. *PLoS Medicine*, *10*(11), e1001547.

Finch, K. C., Snook, K. R., Duke, C. H., Fu, K. W., Tse, Z. T. H., Adhikari, A., & Fung, I. C. H. (2016). Public health implications of social media use during natural disasters, environmental disasters, and other environmental concerns. *Natural Hazards*, *83*(1), 729-760.

Freifeld, C. C., Brownstein, J. S., Menone, C. M., Bao, W., Filice, R., Kass-Hout, T., & Dasgupta, N. (2014). Digital drug safety surveillance: monitoring pharmaceutical products in Twitter. *Drug Safety*, *37*(5), 343-350.

Freud, S. (1957). Mourning and melancholia. In *The Standard Edition of the Complete Psychological Works of Sigmund Freud, Volume XIV (1914-1916): On the History of the Psycho-Analytic Movement, Papers on Metapsychology and Other Works* (pp. 237-258). Hogarth press and the Institute of Psychoanalysis.

Fuchs, C., Hofkirchner, W., Schafranek, M., Raffl, C., Sandoval, M., & Bichler, R. (2010). Theoretical foundations of the web: cognition, communication, and co-operation. towards an understanding of Web 1.0, 2.0, 3.0. *Future Internet*, *2*(1), 41-59.

Gabarron, E., Dorronzoro, E., Rivera-Romero, O., & Wynn, R. (2019). Diabetes on Twitter: a sentiment analysis. *Journal of Diabetes Science and Technology*, *13*(3), 439-444.

Gabilondo, A., Rojas-Farreras, S., Vilagut, G., Haro, J. M., Fernández, A., Pinto-Meza, A., & Alonso, J. (2010). Epidemiology of major depressive episode in a southern European country: results from the ESEMeD-Spain project. *Journal of Affective Disorders, 120*(1-3), 76-85.

Galderisi, S., Heinz, A., Kastrup, M., Beezhold, J., & Sartorius, N. (2015). Toward a new definition of mental health. *World psychiatry: official journal of the World Psychiatric Association (WPA)*, *14*(2), 231-233.

Galderisi, S., Heinz, A., Kastrup, M., Beezhold, J., & Sartorius, N. A proposed new definition of mental health. (2017). *Psychiatria Hungarica*, *51*(3), 407-411.

GBD 2015 DALYs and HALE Collaborators (2016). Global, regional, and national disability-adjusted life-years (DALYs) for 315 diseases and injuries and healthy life expectancy (HALE), 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet*, *388*(10053), 1603-1658.

GBD 2016 Disease and Injury Incidence and Prevalence Collaborators. (2017). Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet*, *390*(10100), 1211-1259.

GBD 2017 Disease and Injury Incidence and Prevalence Collaborators (2018). Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet*, *392*(10159), 1789-1858.

Gilbody, S., House, A., & Sheldon, T. (2005). Screening and case finding instruments for depression. *Cochrane Database of Systematic Reviews*, (4), CD002792.

Gittelman, S., Lange, V., Gotway Crawford, C. A., Okoro, C. A., Lieb, E., Dhingra, S. S., & Trimarchi, E. (2015). A new source of data for public health surveillance: Facebook likes. *Journal of Medical Internet Research*, *17*(4), e98.

Glenn, T., & Monteith, S. (2014). New measures of mental state and behavior based on data collected from sensors, smartphones, and the Internet. *Current Psychiatry Reports*, *16*(12), 523.

Global Burden of Disease Study 2013 Collaborators (2015). Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet*, *386*(9995), 743-800.

Gruber, T. (2007). Collective knowledge systems: where the social web meets the semantic web. *Journal of Web Semantics*: Science, Services, *a6*(1):4-13.

Gureje, O. (2011). Dysthymia in a cross-cultural perspective. *Current Opinion in Psychiatry, 24*(1), 67-71.

Haro, J. M., Palacín, C., Vilagut, G., Martínez, M., Bernal, M., Luque, I., Codony, M., Dolz, M., Alonso, J & Grupo ESEMeD-España (2006). Prevalencia de los trastornos mentales y factores asociados: resultados del estudio ESEMeD-España. *Medicina clínica, 126*(12), 445-451.

Härter, M., Baumeister, H., Reuter, K., Jacobi, F., Höfler, M., Bengel, J., & Wittchen, H. U. (2007). Increased 12-month prevalence rates of mental disorders in patients with chronic somatic diseases. *Psychotherapy and Psychosomatics*, *76*(6), 354-360.

Hasan, H., & Pfaff, C. (2006). The Wiki: an environment to revolutionize employees' interaction with corporate knowledge. *Proceedings of the International Conference on Advances in Computer-Human Interactions* (pp. 377-380). ACM Press.

Helliwell, J. F., Layard, R., & Sachs, J. (2012). World happiness report 2012. UN Sustainable Development Solutions Network. https://happiness-report.s3.amazonaws.com/2020/WHR20.pdf

Hoffman, D. L., Novak, T. P., & Stein, R. (2013). *The digital consumer*. In R. W. Belk and R. Llamas (Eds.), *The Routledge Companion to Digital Consumption Routledge.* (pp. 28-38).

Hswen, Y., Naslund, J. A., Brownstein, J. S., & Hawkins, J. B. (2018). Online communication about depression and anxiety among Twitter users with schizophrenia: preliminary findings to inform a digital phenotype using social media. *The Psychiatric Quarterly, 89*(3), 569-580.

Hu, T. (2004). An international review of the economic costs of mental illness. *World Bank Working Paper 31.*

Huber, M., Knottnerus, J. A., Green, L., van der Horst, H., Jadad, A. R., Kromhout, D., Leonard, B., Lorig, K., Loureiro, M.I., van der Meer, J.W. & Schnabel, P. (2011). How should we define health? *British Medical Journal*, *343*, d4163.

Huguet, A., Rao, S., McGrath, P. J., Wozney, L., Wheaton, M., Conrod, J., & Rozario, S. (2016). A Systematic Review of Cognitive Behavioral Therapy and Behavioral Activation Apps for Depression. *PloS one*, *11*(5), e0154248.

Huppert, F.A., & So, T.T. (2013). Flourishing across Europe: Application of a new conceptual framework for defining well-being. *Social Indicators Research*, *110*(3), 837-61.

Hyman, S., Parikh, R., Collins, P.Y. & Patel, V. (2016). Adult Mental Disorders. In V. Patel et. al. (Eds.), *Mental, Neurological, and Substance Use Disorders: Disease Control Priorities,* (3rd ed., Vol. 4, pp. 67-86). Washington, DC: The World Bank.

Ilieva, R.T. & McPhearson, T. (2018). Social-media data for urban sustainability. *Nature Sustainability*, *1*(10), 553-565.

Insel, T.R. (2018). Digital phenotyping: a global tool for psychiatry. *World Psychiatry*, *17*(3), 276-277.

Instituto Nacional de Estadística. (2018). *Encuesta nacional de salud 2017*. https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Est adistica_C&cid=1254736176783&menu=resultados&idp=1254735573175#!tabs-1254736195650

Isometsä, E. (2014). Suicidal behaviour in mood disorders-who, when, and why? *The Canadian Journal of Psychiatry*, *59*(3), 120-130.

Jacobs, B. L., Van Praag, H., & Gage, F. H. (2000). Adult brain neurogenesis and psychiatry: a novel theory of depression. *Molecular Psychiatry*, *5*(3), 262-269.

Jain, S. H., Powers, B. W., Hawkins, J. B., & Brownstein, J. S. (2015). The digital phenotype. *Nature Biotechnology*, *33*(5), 462-463.

Justicia, A., Elices, M., Cebria, A. I., Palao, D. J., Gorosabel, J., Puigdemont, D., de Diego-Adeliño, J., Gabilondo, A., Iruin, A., Hegerl, U., & Pérez, V. (2017). Rationale and methods of the iFightDepression study: A double-blind, randomized controlled trial evaluating the efficacy of an internet-based self-management tool for moderate to mild depression. *BMC psychiatry*, *17*(1), 143.

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of social media. *Business Horizons*, *53*(1), 59-68.

Kendler, K. S. (2009). An historical framework for psychiatric nosology. *Psychological Medicine, 39*, 1-7.

Kessler, R.C., Chiu, W.T., Demler, O., Merikangas, K.R. & Walters, E.E. (2005). Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, *62*, 617-627.

Keyes, C.L. (2005). Mental illness and/or mental health? Investigating axioms of the complete state model of health. *Journal of Consulting and Clinical Psychology*, *73*(3), 539.

Keyes, C.L. (2014). Mental Health as a Complete State: How the salutogenic perspective completes the picture. In G.F. Bauer & O. Hämmig (Eds.), *Bridging occupational, organizational and public health: A transdisciplinary approach* (pp. 179-192). Springer.

Klein, A., Sarker, A., Rouhizadeh, M., O'Connor, K., & Gonzalez. G. (2017). Detecting personal medication intake in Twitter: an annotated corpus and baseline classification system. *Proceedings of the 16th Biomedical Natural Language Processing (BioNLP)* (pp 136-142). ACL. https://www.aclweb.org/anth(2017)

Kleinman, A. (2009). Global mental health: a failure of humanity. *Lancet*, *374*(9690), 603-604.

Kohn, R., Saxena, S., Levav, I., & Saraceno, B. (2004). The treatment gap in mental health care. *Bulletin of the World Health Organization*, *82*, 858-866.

Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, *110*(15), 5802-5805.

Kraus, C., Kadriu, B., Lanzenberger, R., Zarate Jr, C. A., & Kasper, S. (2019). Prognosis and improved outcomes in major depression: a review. *Translational Psychiatry*, *9*(1), 1-17.

Kriston, L., & von Wolff, A. (2011). Not as golden as standards should be: interpretation of the Hamilton Rating Scale for Depression. *Journal of Affective Disorders*, *128*(1-2), 175-177.

Krueger, R. F. (1999). The structure of common mental disorders. *Archives of General Psychiatry*, *56*(10), 921-926.

Lader, M. (1981). The clinical assessment of depression. *British Journal of Clinical Pharmacology*, *1*, 5-14.

Lardon, J., Bellet, F., Aboukhamis, R., Asfari, H., Souvignet, J., Jaulent, M. C., Beyens, M. N., Lillo-LeLouët, A., & Bousquet, C. (2018). Evaluating Twitter as a complementary data source for pharmacovigilance. *Expert Opinion on Drug Safety*, *17*(8), 763-774.

Lee, E. J., Kim, J. B., Shin, I. H., Lim, K. H., Lee, S. H., Cho, G. A., Sung, H. M., Jung, S. W., Zmimmerman, M., & Lee, Y. (2010). Current use of depression rating scales in mental health setting. *Psychiatry Investigation*, *7*(3), 170-176.

Leis, A., Mayer, M. A., Torres Niño, J., Rodríguez-González, A., Suelves, J. M., & Armayones, M. (2013). Grupos sobre alimentación saludable en Facebook: características y contenidos [Healthy eating support groups on Facebook: content and features]. *Gaceta sanitaria*, *27*(4), 355-357.

Lépine, J. P., & Briley, M. (2011). The increasing burden of depression. *Neuropsychiatric Disease and Treatment*, *7*(Suppl 1), 3-7.

Lewinsohn, P. M. (1974). A behavioral approach to depression. In R. J. Friedman, & M. M. Katz (Eds.), *The psychology of depression: Contemporary theory and research* (pp. 157-178). John Wiley & Sons.

Lewinsohn, P. M., Rohde, P., Seeley, J.R., Klein, D.N. & Gotlib, I.H. (2000). Natural course of adolescent major depressive disorder in a community sample: predictors of recurrence in young adults. *American Journal of Psychiatry, 157*(10), 1584-91.

Liu, J., Weitzman, E. R., & Chunara, R. (2017). Assessing behavioral stages from social media data. *Proceedings of the Conference on Computer-Supported Cooperative Work* (pp. 1320-1333).

Lupiáñez-Villanueva, F., Mayer, M. A., & Torrent, J. (2009). Opportunities and challenges of Web 2.0 within the health care systems: an empirical exploration. *Informatics for Health & Social Care*, *34*(3), 117-126.

Mahata, D., Friedrichs, J., Ratn, R., & Jiang, J. (2018). Did you take the pill? Detecting personal intake of medicine from Twitter. *IEEE Intelligent System, 33*(4):87-95.

Manwell, L. A., Barbic, S. P., Roberts, K., Durisko, Z., Lee, C., Ware, E., & McKenzie, K. (2015). What is mental health? Evidence towards a new definition from a mixed methods multidisciplinary international survey. *British Medical Journal Open*, *5*(6), e007079.

Marcus, M., Yasamy, M.T., van Ommeren, M., Chisholm, D., & Saxena, S. (2012). *Depression: A Global Public Health Concern*. WHO Department of Mental Health and Substance Abuse. World Health Organization.

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

Martin-Vazquez, M. J. (2016). Adherence to antidepressants: a review of the literature. *Neuropsychiatry, 6*(5), 236-241.

Mayer, M. A., Fernández-Luque, L., & Leis, A. (2016). Big Data for health through social media. In S. Syed-Abdul, E. Gabarron & A. Lau (Eds.), *Participatory Health Through Social Media* (pp. 67-82). Academic Press. Elsevier Inc.

Meng, L., Chen, D., Yang, Y., Zheng, Y., & Hui, R. (2012). Depression increases the risk of hypertension incidence: a meta-analysis of prospective cohort studies. *Journal of Hypertension*, *30*(5), 842-851.

Meskó, B. (2013). The role of Twitter and microblogging in medicine. *Social media in clinical practice* (pp. 71-78). Springer.

Mnookin, S. (2016). *Out of the shadows: making mental health a global development priority*. Washington, D.C.: World Bank Group. http://documents1.worldbank.org/curated/en/270131468187759113/pdf/105052-WP-PUBLIC-wb-background-paper.pdf

Montgomery, S. A. (1980). Measurement of serum drug levels int he assessment of antidepressants. *British Journal of Clinical Pharmacology, 10*, 41-416.

Morales, M. R. (2018). *Multimodal Depression Detection: An Investigation of Features and Fusion Techniques for Automated Systems.* CUNY Academic Works.

Morris, D. W., Rush, A. J., Jain, S., Fava, M., Wisniewski, S. R., Balasubramani, G. K., Khan, A. Y., & Trivedi, M. H. (2007). Diurnal mood variation in outpatients with major depressive disorder: implications for DSM-V from an analysis of the sequenced treatment alternatives to relieve depression Study data. *The Journal of Clinical Psychiatry*, *68*(9), 1339-1347.

Mowery, D., Smith, H., Cheney, T., Stoddard, G., Coppersmith, G., Bryan, C., & Conway, M. (2017). Understanding depressive symptoms and psychosocial stressors on Twitter: a corpus-based study. *Journal of Medical Internet Research*, *19*(2), e48.

Murray, C. J., & Lopez, A. D. (2013). Measuring the global burden of disease. *New England Journal of Medicine*, *369*(5), 448-457.

Nambisan, P., Luo, Z., Kapoor, A., Patrick, T. B., & Cisler, R. A. (2015). Social media, big data, and public health informatics: ruminating behavior of depression revealed through twitter. *Proceedings of the 2015 48th Hawaii International Conference on System Sciences* (pp. 2906-2913). IEEE.

Naslund, J. A., Aschbrenner, K. A., McHugo, G. J., Unützer, J., Marsch, L. A., & Bartels, S. J. (2019). Exploring opportunities to support mental health care using social media: a survey of social media users with mental illness. *Early Intervention in Psychiatry*, *13*(3), 405-413.

Naslund, J. A., Gonsalves, P. P., Gruebner, O., Pendse, S. R., Smith, S. L., Sharma, A., & Raviola, G. (2019). Digital innovations for global mental health: opportunities for data science, task sharing, and early intervention. *Current Treatment Options in Psychiatry*, *6*(4), 337-351.

National Institute for Health and Care Excellence (NICE). (2018). Depression in adults: recognition and management. Clinical guideline [CG90]. https://www.nice.org.uk/guidance/cg90/resources/depression-in-adults-recognition-and-management-pdf-975742638037

National Library of Medicine. (2012). Social Media. In *Medical Subject Headings*. Retrieved October 30, 2020, from https://www.ncbi.nlm.nih.gov/mesh/?term=web+2.0

Neacsiu, A. D., Rompogren, J., Eberle, J. W., & McMahon, K. (2018). Changes in problematic anger, shame, and disgust in anxious and depressed adults undergoing treatment for emotion dysregulation. *Behavior Therapy, 49*(3), 344-359.

NRC-Canada at SMM4H shared task: classifying tweets mentioning adverse drug reactions and medication intake. *Proceedings of the 2nd Social Media Mining for Health Applications Workshop co-located with the American Medical Informatics Association Annual Symposium* (AMIA)(pp.1-11). http://ceur-ws.org/Vol-1996/paper1.pdf

Nguyen, T., O'Dea, B., Larsen, M., Phung, D., Venkatesh, S., & Christensen, H. (2015). Using linguistic and topic analysis to classify sub-groups of online depression communities. *Multimedia Tools and Applications*, *76*(8), 10653-10676.

Nikfarjam, A., Ransohoff, J. D., Callahan, A., Jones, E., Loew, B., Kwong, B. Y., Sarin, K. Y., & Shah, N. H. (2019). Early detection of adverse drug reactions in social health networks: a natural language processing pipeline for signal detection. *JMIR public health and surveillance, 5*(2), e11264.

Nikfarjam, A., Sarker, A., O'connor, K., Ginn, R., & Gonzalez, G. (2015). Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features. *Journal of the American Medical Informatics Association*, *22*(3), 671-681.

NLTK Project. (2020, April 13). Natural Language Tool Kit. https://www.nltk.org/api/nltk.tokenize.html

Obar, J. A., & Wildman, S. S. (2015). Social media definition and the governance challenge-an introduction to the special issue. *Telecommunications Policy, 39(9), 745-750.*

O'Connor, K., Pimpalkhute, P., Nikfarjam, A., Ginn, R., Smith, K. L., & Gonzalez, G. (2014). Pharmacovigilance on twitter? Mining tweets for adverse drug reactions. *AMIA Annual Symposium proceedings* (pp. 924-933).

O'Reilly, T. (2005). What is Web 2.0: design patterns and business models for the next generation of software. https://www.oreilly.com/pub/a/web2/archive/what-is-web-20.htm

Padró, L., & Stanilovsky, E. (2012). FreeLing 3.0: Towards Wider Multilinguality. *Proceedings of the Eighth International Conference on Language Resources and Evaluation* (pp. 2473-2479). ELRA.

Painuly, N., Sharan, P., & Mattoo, S. K. (2005). Relationship of anger and anger attacks with depression. European Archives of Psychiatry and Clinical Neuroscience, *255*(4), 215-222.

Park, M., Cha, C., & Cha, M. (2012). Depressive moods of users portrayed in Twitter. *Proceedings of the ACM SIGKDD Workshop on Health Informatics*, (pp.1-8). AAAI Press.

Paul, M.J., & Dredze, M. (2011). You are what you tweet: analyzing Twitter for public health. *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media* (pp. 265-272). AAAI Press.

Paul, M. J., & Dredze, M. (2014). Discovering health topics in social media using topic models. *PLoS One*, *9*(8), e103408.

Patel, V., Lund, C., Heatherill, S., Plagerson, S., Corrigal, J., Funk, M., & Flisher, A. (2009). Social determinants of mental disorders. *Priority public health conditions: from learning to action on social determinants of health. Geneva: World Health Organization*.

Patel, V. & Saxena, S. (2014). Transforming lives, enhancing communities - innovations in global mental health. *New England Journal of Medicine, 370*(6), 498-501.

Patel, V., Chisholm, D., Parikh, R., Charlson, F. J., Degenhardt, L., Dua, T., Ferrari, A. J., Hyman, S., Laxminarayan, R., Levin, C., Lund, C., Medina-Mora, M. E., Petersen, I., Scott, J. G., Shidhaye, R., Vijayakumar, L., Thornicroft, G., & Whiteford, H. A., on behalf of the DCP MNS authors group (2016). Global priorities for addressing the burden of mental, neurological, and substance use disorders. In V. Patel et. al. (Eds.), *Mental, Neurological, and Substance Use Disorders: Disease Control Priorities,* (3rd ed., Vol. 4, pp. 1-27). Washington, DC: The World Bank.

Patel, V., Saxena, S., Lund, C., Thornicroft, G., Baingana, F., Bolton, P., Chisholm, D., Collins, P. Y., Cooper, J. L., Eaton, J., Herrman, H., Herzallah, M. M., Huang, Y., Jordans, M., Kleinman, A., Medina-Mora, M. E., Morgan, E., Niaz, U., Omigbodun, O., Prince, M., … UnÜtzer, J. (2018). The Lancet Commission on global mental health and sustainable development. *Lancet*, *392*(10157), 1553-1598.

Pathare, S., Brazinova, A., & Levav, I. (2018). Care gap: a comprehensive measure to quantify unmet needs in mental health. *Epidemiology and Psychiatric Sciences*, *27*(5), 463-467.

Pavalanathan, U., & De Choudhury, M. (2015). Identity management and mental health discourse in social media. *Proceedings of the 24th International Conference on World Wide Web* (pp. 315-321). ACM.

Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology*, *54*(1), 547-577.

Pérez, V., Salavert, A., Espadaler, J., Tuson, M., Saiz-Ruiz, J., Sáez-Navarro, C., Bobes, J., Baca-García, E., Vieta, E., Olivares, J. M., Rodríguez-Jiménez, R., Villagrán, J. M., Gascón, J., Cañete-Crespillo, J., Solé, M., Saiz, P. A., Ibáñez, A., de Diego-Adeliño, J., AB-GEN Collaborative Group & Menchón, J. M. (2017). Efficacy of prospective pharmacogenetic testing in the treatment of major depressive disorder: results of a randomized, double-blind clinical trial. *BMC Psychiatry*, *17*(1), 1-13.

Pérez-Rosas, V., Banea, C., & Mihalcea, R. Learning Sentiment Lexicons in Spanish. (2012). *Proceedings of the Eighth International Conference on Language Resources and Evaluation* (pp. 3077-3081). ELRA. http://www.lrec-conf.org/proceedings/lrec2012/pdf/1081_Paper.pdf

Pierce, C. E., Bouri, K., Pamer, C., Proestel, S., Rodriguez, H. W., Van Le, H., Freifeld , C. C., Brownstein, J. S., Walderhaug, M., Edwards, I. R., & Dasgupta, N. (2017). Evaluation of Facebook and Twitter monitoring to detect safety signals for medical products: an analysis of recent FDA safety alerts. *Drug Safety*, *40*(4), 317-331.

Prieto, V. M., Matos, S., Alvarez, M., Cacheda, F., & Oliveira, J. L. (2014). Twitter: a good place to detect health conditions. *PloS one*, *9*(1), e86191.

Pyszczynski, T., & Greenberg, J. (1987). Self-regulatory perseveration and the depressive self-focusing style: a self-awareness theory of reactive depression. *Psychological Bulletin*, *102*(1), 122.

Radzikowski, J., Stefanidis, A., Jacobsen, K. H., Croitoru, A., Crooks, A., & Delamater, P. L. (2016). The measles vaccination narrative in Twitter: a quantitative analysis. *JMIR Public Health and Surveillance*, *2*(1), e1.

Ramirez-Esparza, N., Chung, C. K., Kacewicz, E., & Pennebaker, J. W. (2008, March). The psychology of word use in depression forums in english and in spanish: texting two text analytic approaches. *Proceedings of the 2nd International Conference on Weblogs and Social Medi*a (pp. 102-108). AAAI Press. https://www.aaai.org/Papers/ICWSM/2008/ICWSM08-020.pdf

Reece, A. G., & Danforth, C. M. (2017). Instagram photos reveal predictive markers of depression. *EPJ Data Science*, *6*(1), 1-12.

Reece, A. G., Reagan, A. J., Lix, K. L., Dodds, P. S., Danforth, C. M., & Langer, E. J. (2017). Forecasting the onset and course of mental illness with Twitter data. *Scientific Reports*, *7*(1), 1-11.

Reed, G. M., Correia, J., Esparza, P., Saxena, S., & Maj, M. (2011). The WPA-WHO global survey of psychiatrists' attitudes towards mental disorders classification. *World Psychiatry*, *10*(2), 118-131.

Rehm, J., & Shield, K.D. (2019). Global Burden of Disease and the Impact of Mental and Addictive Disorders. *Current Psychiatry Reports*, *21*(2), 10.

Ricard, B. J., Marsch, L. A., Crosier, B., & Hassanpour, S. (2018). Exploring the utility of community-generated social media content for detecting depression: an analytical study on Instagram. *Journal of Medical Internet Research*, *20*(12), e11817.

Royal College of Psychiatrist. (2019). Position statement on antidepressants and depression.PS04/19. https://www.rcpsych.ac.uk/docs/default-source/improving-care/better-mh-policy/position-statements/ps04_19---antidepressants-and-depression.pdf?sfvrsn=ddea9473_5

Rozatkar, A. R., & Gupta, N. (2016). Global burden of psychiatric disorders: has it increased? *Indian Journal of Social Psychiatry*, *32*(3), 249-250.

Rude, S., Gortner, E. M., & Pennebaker, J. (2004). Language use of depressed and depression-vulnerable college students. *Cognition & Emotion*, *18*(8), 1121-1133.

Saeb, S., Zhang, M., Karr, C. J., Schueller, S. M., Corden, M. E., Kording, K. P., & Mohr, D. C. (2015). Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study. *Journal of Medical Internet Research*, *17*(7), e175.

Sadah, S. A., Shahbazi, M., Wiley, M. T., & Hristidis, V. (2015). A study of the demographics of web-based health-related social media users. *Journal of Medical Internet Research, 17*(8), e194.

Saha, K., Sugar, B., Torous, J., Abrahao, B., Kıcıman, E., & De Choudhury, M. (2019). A Social media study on the effects of psychiatric medication use. *Proceedings of the thirteenth International AAAI Conference on Weblogs and Social Media* (pp. 440-451). AAAI.

Shadrina, M., Bondarenko, E. A., & Slominsky, P. A. (2018). Genetics Factors in Major Depression Disease. *Frontiers in psychiatry*, *9*, 334.

Salathé, M. (2016). Digital pharmacovigilance and disease surveillance: combining traditional and big-data systems for better public health. *The Journal of Infectious Diseases*, *214*(4), S399-S403.

Salathé, M. (2018). Digital epidemiology: what is it, and where is it going? *Life Sciences, Society and Policy*, *14*(1).

San Too, L., Spittal, M. J., Bugeja, L., Reifels, L., Butterworth, P., & Pirkis, J. (2019). The association between mental disorders and suicide: a systematic review and meta-analysis of record linkage studies. *Journal of Affective Disorders*, *259*, 302-313

Sarker, A., Ginn, R., Nikfarjam, A., O'Connor, K., Smith, K., Jayaraman, S., Upadhaya, T., & Gonzalez, G. (2015). Utilizing social media data for pharmacovigilance: a review. *Journal of Biomedical Informatics*, *54*, 202-212.

Sartorius, N. (2017). Comorbidity of Mental and Physical Disorders: a major challenge for medicine in the 21st century. *European Psychiatry*, *41*(S1), S9-S9.

Saxena, S., Thornicroft, G., Knapp, M., & Whiteford, H. (2007). Resources for mental health: scarcity, inequity, and inefficiency. *Lancet*, *370*(9589), 878-889.

Schwartz, S., Schultz, S., Reider, A., & Saunders, E. F. (2016). Daily mood monitoring of symptoms using smartphones in bipolar disorder: a pilot study assessing the feasibility of ecological momentary assessment. *Journal of Affective Disorders*, *191*, 88-93.

Seabrook, E. M., Kern, M. L., Fulcher, B. D., & Rickard, N. S. (2018). Predicting depression from language-based emotion dynamics: longitudinal analysis of Facebook and Twitter status updates. *Journal of Medical Internet Research, 20*(5), e168.

Segura-Bedmar, I., Martínez, P., Revert, R., & Moreno-Schneider, J. (2015). Exploring Spanish health social media for detecting drug effects. *BMC medical informatics and decision making, 15*(Suppl 2), S6.

Sharp, L. K., & Lipsky, M. S. (2002). Screening for depression across the lifespan: a review of measures for use in primary care settings. *American Family Physician*, *66*(6), 1001-1008.

Sidorov, G., Miranda-Jiménez, S., Viveros-Jiménez, F., Gelbukh, A., Castro-Sánchez, N.A., Castillo, F., Velásquez, F., Díaz-Rangel, I., Suárez-Guerra, S., Treviño, A., & Gordon, J. (2012). Empirical study of opinion mining in Spanish tweets. *Proceedings of the 11th Mexican International Conference on Artificial Intelligence* (pp. 1-4). IEEE Computer Society.

Sonnenschein, A. R., Hofmann, S. G., Ziegelmayer, T., & Lutz, W. (2018). Linguistic analysis of patients with mood and anxiety disorders during cognitive behavioral therapy. *Cognitive Behaviour Therapy*, *47*(4), 315-327.

Soreni, N., Cameron, D. H., Streiner, D. L., Rowa, K., & McCabe, R. E. (2019). Seasonality patterns of internet searches on mental health: exploratory infodemiology study. *JMIR Mental Health, 6*(4), e12974.

Stefanidis, A., Vraga, E., Lamprianidis, G., Radzikowski, J., Delamater, P. L., Jacobsen, K. H., Pfoser, D., Croitoru, A., & Crooks, A. (2017). Zika in Twitter: temporal variations of locations, actors, and concepts. *JMIR Public Health and Surveillance*, *3*(2), e22.

Stein, D., Phillips, K., Bolton, D., Fulford, K., Sadler, J., & Kendler, K. (2010). What is a mental/psychiatric disorder? From DSM-IV to DSM-V. *Psychological Medicine, 40*(11), 1759-1765.

Stein, D. (2013). What is a mental disorder? A perspective from cognitive-affective science. *Canadian Journal of Psychiatry*, *58*(12), 656-662.

Stenmark, D. (2008). Web 2.0 in the business environment: the new Intranet or a passing hype? *Proceedings of the 16th European Conference on Information Systems* (pp. 2064-2075).

Stirman, S. W., & Pennebaker, J. W. (2001). Word use in the poetry of suicidal and nonsuicidal poets. *Psychosomatic Medicine*, *63*(4), 517-522.

Strine, T. W., Mokdad, A. H., Balluz, L. S., Gonzalez, O., Crider, R., Berry, J. T., & Kroenke, K. (2008). Depression and anxiety in the United States: findings from the 2006 behavioral risk factor surveillance system. *Psychiatric Services (Washington, D.C.)*, *59*(12), 1383-1390.

Subirats, L., Reguera, N., Bañón, A. M., Gómez-Zúñiga, B., Minguillón, J., & Armayones, M. (2018). Mining Facebook Data of People with Rare Diseases: A Content-Based and Temporal Analysis. *International Journal of Environmental Research and Public Health*, *15*(9), 1877. https://doi.org/10.3390/ijerph15091877

Suldo, S.M., & Shaffer, E. J. (2008). Looking beyond psychopathology: the dual-factor model of mental health in youth. *School Psychology Review*, *37*(1), 52-68.

Surian, D., Nguyen, D. Q., Kennedy, G., Johnson, M., Coiera, E., & Dunn, A. G. (2016). Characterizing Twitter discussions about HPV vaccines using topic modeling and community detection. *Journal of Medical Internet Research*, *18*(8), e232.

Tadesse, M. M., Lin, H., Xu, B., & Yang, L. (2019). Detection of depression-related posts in reddit social media forum. *IEEE Access*, *7*, 44883-44893.

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, *29*(1), 24-54.

Thackeray, R., Burton, S. H., Giraud-Carrier, C., Rollins, S., & Draper, C. R. (2013). Using Twitter for breast cancer prevention: an analysis of breast cancer awareness month. *BMC Cancer*, *13*(1), 508.

Thompson, L. A., Dawson, K., Ferdig, R., Black, E. W., Boyer, J., Coutts, J., & Black, N. P. (2008). The intersection of online social networking with medical professionalism. *Journal of General Internal Medicine*, *23*(7), 954-957.

Trautmann, S., Rehm, J., & Wittchen, H. U. (2016). The economic costs of mental disorders: Do our societies react appropriately to the burden of mental disorders? *EMBO Reports*, *17*(9), 1245-1249.

Treem, J. W., & Leonardi, P. M. (2013). Social media use in organizations: exploring the affordances of visibility, editability, persistence, and association. *Annals of the International Communication Association*, *36*(1), 143-189.

Tredinnick, L. (2006). Web 2.0 and business. *Business Information Review*, *23*(4), 228-234.

Üstün, T.B., & Roger, Ho. (2017). Classification of Mental Disorders: Principles and Concepts. In S. R. Quah (Ed.). *International Encyclopedia of Public Health* (2nd ed., pp. 51-57). Academic Press.

Vaillant G. E. (2012). Positive mental health: is there a cross-cultural definition? *World psychiatry: official journal of the World Psychiatric Association (WPA)*, *11*(2), 93-99.

van Westrhenen, R., Aitchison, K. J., Ingelman-Sundberg, M., & Jukić, M. M. (2020). Pharmacogenomics of antidepressant and antipsychotic treatment: how far have we got and where are we going?. *Frontiers in Psychiatry*, *11*.

Vigo, D., Thornicroft, G., & Atun, R. (2016). Estimating the true global burden of mental illness. *The Lancet Psychiatry*, *3*(2), 171-178.

Vigo, D. V., Kestel, D., Pendakur, K., Thornicroft, G., & Atun, R. (2019). Disease burden and government spending on mental, neurological, and substance use disorders, and self-harm: cross-sectional, ecological study of health system response in the Americas. *The Lancet Public Health*, *4*(2), e89-e96.

Walker, E. R., McGee, R. E., & Druss, B. G. (2015). Mortality in mental disorders and global disease burden implications: a systematic review and meta-analysis. *JAMA Psychiatry*, *72*(4), 334-341.

Wang, P. S., Aguilar-Gaxiola, S., Alonso, J., Angermeyer, M. C., Borges, G., Bromet, E. J., Bruffaerts, R., de Girolamo, G., de Graaf, R., Gureje, O., Haro, J. M., Karam, E. G., Kessler, R. C., Kovess, V., Lane, M. C., Lee, S., Levinson, D., Ono, Y., Petukhova, M., Posada-Villa, J., … Wells, J. E. (2007). Use of mental health services for anxiety, mood, and substance disorders in 17 countries in the WHO world mental health surveys. *The Lancet*, *370*(9590), 841-850.

Wang, T., Brede, M., Ianni, A., & Mentzakis, E. (2018). Social interactions in online eating disorder communities: A network perspective. *PloS one, 13*(7), e0200800.

Weinberg, B. D., & Pehlivan, E. (2011). Social spending: managing the social media mix. *Business Horizons*, *54*(3), 275-282.

Weintraub, W. (1981). *Verbal behavior: adaptation and psychopathology*. New York: Springer Publishing Company.

Wenzel, M., Kubiak, T., & Ebner-Priemer, U. W. (2016). Ambulatory assessment as a means of longitudinal phenotypes characterization in psychiatric disorders. *Neuroscience Research*, *102*, 13-2.

Whiteford, H. A., Degenhardt, L., Rehm, J., Baxter, A. J., Ferrari, A. J., Erskine, H. E., Charlson, F. J., Norman, R. E., Flaxman, A. D., Johns, N., Burstein, R., Murray, C. J., & Vos, T. (2013). Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. *Lancet*, *382*(9904), 1575-1586.

Whiteford, H. A., Ferrari, A. J., Degenhardt, L., Feigin, V., & Vos, T. (2015). The Global Burden of Mental, Neurological and Substance Use Disorders: An Analysis from the Global Burden of Disease Study 2010. *PLoS One*, *10*(2), e0116820.

Whiteford, H., Ferrari, A., & Degenhardt, L. (2016). Global burden of disease studies: implications for mental and substance use disorders. *Health Affairs*, *35*(6), 1114-1120.

Wilson, M.L., Ali, S., & Valstar, M.F. (2014). Finding information about mental health in microblogging platforms: a case study of depression. *Proceedings of the 5th Information Interaction in Context Symposium* (pp. 8-17). ACM.

Wittchen, H. U., Jacobi, F., Rehm, J., Gustavsson, A., Svensson, M., Jönsson, B., Olesen, J., Allgulander, C., Alonso, J., Faravelli, C., Fratiglioni, L., Jennum, P., Lieb, R., Maercker, A., van Os, J., Preisig, M., Salvador-Carulla, L., Simon, R., & Steinhausen, H. C. (2011). The size and burden of mental disorders and other disorders of the brain in Europe 2010. *European Neuropsychopharmacology*, *21*(9), 655-679.

Wongkoblap, A., Vadillo, M. A., & Curcin, V. (2017). Researching mental health disorders in the era of social media: systematic review. *Journal of Medical Internet Research*, *19*(6), e228.

World Health Organization. (2013). *Mental Health Action Plan 2013-2020*. http://apps.who.int/iris/bitstream/10665/89966/1/9789241506021_eng.pdf?ua=1

World Health Organization. (2014). *Global Health Estimates 2014 Summary Tables: YLD by Cause, Age and Sex, 2000-2012*. http://www.who.int/healthinfo/global_burden_disease/en/

World Health Organization. (2017). *Depression and other common mental disorders: global health estimates*. https://apps.who.int/iris/bitstream/handle/10665/254610/WHW-MSD-MER-2017.2-eng.pdf?sequence=1

World Health Organization. (2018a). *Mental health atlas 2017*. WHO.

World Health Organization. (2018b). *mhGAP operations manual: mental health Gap Action Programme (mhGAP)*. WHO.

World Health Organization. (2019a). Chapter 6. Mental, behavioural and neurodevelopmental disorders. *International statistical classification of diseases and related health problems (11th Revision).* WHO.

World Health Organization. (2019b). *Mental disorders: Key Facts*. Retrieved July 2, 2020, from https://www.who.int/en/news-room/fact-sheets/detail/mental-disorders

World Medical Association (WMA). (2017). *WMA Statement on the professional and ethical use of social media.* https://www.wma.net/policies-post/wma-statement-on-the-professional-and-ethical-use-of-social-media/

World Wide Web Consortium (W3C). (2020). Retrieved September 15, 2020, from https://www.w3.org/

Yom-Tov, E., Johansson-Cox, I., Lampos, V., & Hayward, A. C. (2015). Estimating the secondary attack rate and serial interval of influenza-like illnesses using social media. *Influenza and Other Respiratory Viruses*, *9*(4), 191-199.

Youyou, W., Kosinski, M., & Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences*, *112*(4), 1036-1040.

Zinken, J., Zinken, K., Wilson, J. C., Butler, L., & Skinner, T. (2010). Analysis of syntax and word use to predict successful participation in guided self-help for anxiety and depression. *Psychiatry Research*, *179*(2), 181-186.