System medicine approach for the comorbidity study

Alba Gutiérrez-Sacristán

TESI DOCTORAL UPF / 2017

DIRECTOR DE LA TESI

Dra. Laura Inés Furlong

DEPARTAMENT DE CIÈNCIES EXPERIMENTALS I DE LA SALUT

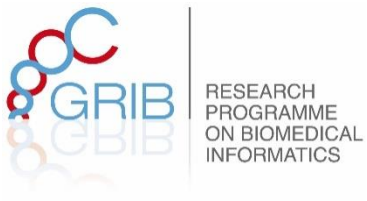
[Pàgina en blanc]

[S'han marcat els apartats de la tesi que haurien d'anar en pàgina imparell o dreta de la publicació (situació de pàgines que ajuda a estructurar i a donar prioritat formal als diferents apartats).

Hi ha apartats on no s'indica res. En aquests cas el doctorand pot triar la situació, en pàgina parell o senar, segons li convingui per al compaginat final.

Es recomana que les pàgines blanques no portin número de foli (es compten, però el número no s'imprimeix)]

The research leading to these results has received support from the Spanish Ministry of Economy and Competitiveness, through the "María de Maeztu" Programme for Units of Excellence in R&D (MDM-2014-0370). ISCIII-FEDER (PI13/00082, CP10/00524), IMI-JU under grants 326 agreements n 115002 (eTOX), n 115191 (Open PHACTS)], n 115372 (EMIF) and n 115735 327 (iPiE), resources of which are composed of financial contribution from the EU-FP7 (FP7/2007- 328 2013) and EFPIA companies in kind contribution, and the EU H2020 Programme 2014-2020 329 under grant agreements no. 634143 (MedBioinformatics) and no. 676559 (Elixir-Excelerate). 330 The Research Programme on Biomedical Informatics (GRIB) is a node of the Spanish National 331 Institute of Bioinformatics (INB).



Para mamá y papá.

Para ti abuelo Helios, mi inspiración.

[Pàgina en blanc]

Acknowledgements

Text

[Pàgina en blanc]

Abstract

Clinical and epidemiological studies show that comorbidity, the coexistence of disorders in a patient, has a great impact in the health status evolution. Therefore, understanding the etiology of comorbidities is key to identify new preventive and therapeutic strategies, and to improve the management of these patients.

In order to harness the power of emerging disease data, this thesis presents a new approach for the exploitation of information contained in clinical health records as well as the identification of comorbidity patterns. Furthermore, by exploiting text mining tools, this thesis also presents PsyGeNET, a manual curated gene-disease database, focus on psychiatric disorders that engages the study of molecular and cellular mechanisms that underpin the disease comorbidities in the psychiatrist field.

The tools developed during this thesis are available to the scientific community and they have been already applied in multiple studies in the biomedical field. They allow transforming clinical information in knowledge that can be analyzed, interpreted by doctors, and applied in order to improve patient's life, a step more toward the personalized medicine.

Resum

Estudios clínicos y epidemiológicos muestran que la comorbilidad, la coexistencia de varias enfermedades en un mismo paciente, tiene un gran impacto en la evolución de su estado de salud. Por lo tanto, entender la etiología de la comorbilidad es clave para identificar nuevas estrategias preventivas y terapéuticas y mejorar el manejo de estos pacientes.

Con el fin de aprovechar las grandes cantidades de información sobre los pacientes y sus enfermedades de las que actualmente disponemos, esta tesis presenta un nuevo enfoque para la explotación de la información contenida en los registros clínicos de salud así como la identificación de patrones de comorbilidad. Además, aplicando herramientas de minería de texto, esta tesis también presenta PsyGeNET, una base de datos curada manualmente por expertos, que contiene asociaciones gen-enfermedad centrada en trastornos psiquiátricos. Esta base de datos permitirá el estudio de mecanismos moleculares y celulares que influirían en las comorbilidades de las enfermedades psiquiátricas.

Las herramientas desarrolladas durante esta tesis están disponibles para la comunidad científica y se han aplicado ya en múltiples estudios en el campo biomédico. Transformar la información clínica en conocimiento que puede ser analizado e interpretado por los médicos y aplicado para mejorar la vida del paciente, supone un paso más hacia la medicina personalizada.

[Pàgina en blanc]

Preface

"*Don't limit your challenges, challenge your limits!*" This quote summarize the last years in my working life. During this PhD every day has been a great challenge, and the main goal has been continue growing without losing motivation.

As bioinformatician with specific interest in biomedical research, during the last years I have explored a wide range of bioinformatic approaches, from molecular dynamics and systems biology, to network medicine. In 2013 I joined to the Integrative Biomedical Informatics (IBI) Group in order to conduct research towards a PhD degree in bioinformatics, under the direction of Dr. Laura I. Furlong. In the undertaken PhD work, my research has mainly focused on the development of bioinformatic approaches to analyse disease comorbidities. Developing a system medicine approach for the exploration of the relationships between diseases at different levels will provide key information to a better understanding of the biologic mechanisms of comorbidities.

At the end of this thesis, and after having exploited clinical data from patient record databases to identify comorbid diseases, I started to analyze comorbidities from the genomic point of view, exploring the molecular and cellular mechanism that substantiate comorbidities, more specifically for psychiatric disorders comorbidities, gaining a broad insight into the disease comorbidities.

This thesis is organized as follows: the challenging task of analyzing clinical health record data and the comorbidity concept will be introduced in Chapter 1. Furthermore, the need of developing new strategies for the analysis of clinical data as well as the necessity of having molecular curated information related to Psychiatric disorders will be discussed and current tools in biomedical field will be described. In Chapter 2, the motivation and objectives of this thesis will be presented. The general methodology to the comorbidity study will be described in Chapter 3, introducing the comoRbidity R package as a tool for analyze clinical health records. Complementary information, applications and results of the comorbidity software and PsyGeNET database will be presented in Chapter 4. In Chapter 5 a discussion of the work conducted in this thesis, together with limitations and future perspectives will be provided. Conclusions will be drawn in Chapter 6. Finally, selected publications, in which this work has been applied, will be listed in the Appendix.

[Pàgina en blanc]

Table of contents

List of figures

List of table

[Pàgina en blanc]

1 Introduction

"Science knows no country, because
knowledge belongs to humanity and
is the torch which illuminates the world"
Louis Pasteur (1822-1895)

1.1 Introduction to the comorbidity

Mr. A. sees his general practitioner because he has one symptom, he has been coughing for few days and has fever. His doctor learnt at medical school what the diseases are that can present with this symptom - pharyngitis, pneumonia, pulmonary cancer -, and his clinical exam and some lab tests will help to make the right diagnosis, pharyngitis. Then he will prescribe some treatment for this specific diagnosis. This is the naïve way to believe how medicine works.

In fact, most of the patients have several symptoms that reflect several diseases occurring at the same time in the same patient. Moreover, some of them are acute, other are chronic. For example, Mr. A may have the co-occurrence of a chronic cardiac disease and an acute pharyngitis and both can be responsible for cough. Lots of new questions come up: "*Is a pulmonary cancer more likely because of the chronic cardiac disease, even if the patient has pharyngitis*? And if yes, *is there a need of specific tests in this specific patient with a cough*? *Which drugs to treat the pharyngitis are not harmful in a patient with a cardiac chronic disease*? *Can the cardiac chronic disease modify the results of the lab*? ...

Patients usually do not present only one disorder, the co-occurrence of bunch of disorders in the same patients are frequently encountered, being the rule rather than an exception, and this come up a lot of issues. Alvan R. Feinstein was the first to emphasize this problematic in his seminal paper in 1970, in which he also coined a term for the co-existing disorders, what he defined as comorbidity.

1.1.1 The concept: origin and discrepancies

Feinstein defined 'comorbidity' as "*any distinct clinical entity that has co-existed or that may occur during the clinical course of a patient who has the index disease under study*" (Feinstein 1970) .

Since then, the concept has evolved in different directions, becoming an issue of concern in clinical care. Nowadays, several different definitions exist, but there is not a consensus one in the same way as there is not a uniform methodology to the

comorbidity study. Many authors assuming that the meaning of the concept is widely understood, used it without providing any definition (Almirall et al. 2013), and when it is provided, the exact meaning or definition varies from one author to another (Table 1).

1.1.2 Comorbidity definitions: literature review

Analyzing a set of articles related to comorbidity, more than 10 different definitions can be found related to disease co-occurrence (Table 1). They mainly differ in considering or not an index disease and also taking or not into account only chronic diseases.

Table 1 General comorbidity' definitions widespread in the literature. The comorbidity definitions are sorted by date, from the most recent one to the first one given by Feinstein in 1970. For each definition, the table contains the author, the data and an approximate number of publications using this definition.

Definition	***Author and date***	***Publications***
Coexistence of two or more long-term conditions in one patient / different systems of the body	(Lawson et al. 2013)	
Coexistence of multiple illnesses of different types	(Starfield & Kinder 2011)	
Co-occurrence of multiple medical conditions within one person without any reference to an index condition	(Bayliss et al. 2008)	
Case where an individual suffers from two or more disease conditions at the same time / within a specific period of time	(Marengoni et al. 2008)	
Coexistence of three or more clinical conditions	(Cesari et al. 2006)	
Coincidence of two or more diseases in a patient	(Mikuls & Saag 2001)	
Co-occurrence of multiple chronic or acute diseases and medical conditi within one person	(Van Den Akker et al. 2001)	
Co-existence of two or more chronic conditions, where one is not necessarily more central than the others	(van den Akker et al. 1996)	
Co-occurrence of multiple (often chronic) diseases or medical conditions within one person	(McGee et al. 1996)	
Coexistence of two or more chronic diseases in the same individual	(Schellevis et al. 1993)	
Co-occurrence of several chronic conditions simultaneously	(Verbrugge et al. 1989)	
Any distinct clinical entity that has co-existed or that may occur during the clinical course of a patient who has the index disease under study	(Feinstein 1970)	

The most frequently used definitions are "*coexistence of two or more chronic diseases in the same individual*" and *"coincidence of two or more diseases in a patient"*. The

Feinstein and Bayliss et al. definitions differ in considering or not an index disease, while Milkus et al. and van den Akker et al. have a different approach in considering chronicity. Nevertheless, all terms and definitions refer to the co-occurrence of multiple diseases (van den Akker et al. 1996; Bonavita & De Simone 2008).

This bibliometric analysis of definitions referring to the presence of co-occurring disorders in one patients evidence the diversity in the medical literature to describe this common situation, and differences in the frequency of their use.

1.1.3 Alternatives to comorbidity concept

Apart from comorbidity, other terms such as 'multimorbidity', 'multipathology', 'morbidity burden', 'multicondition' or 'patient complexity' are used in the same context (Valderas et al. 2009).

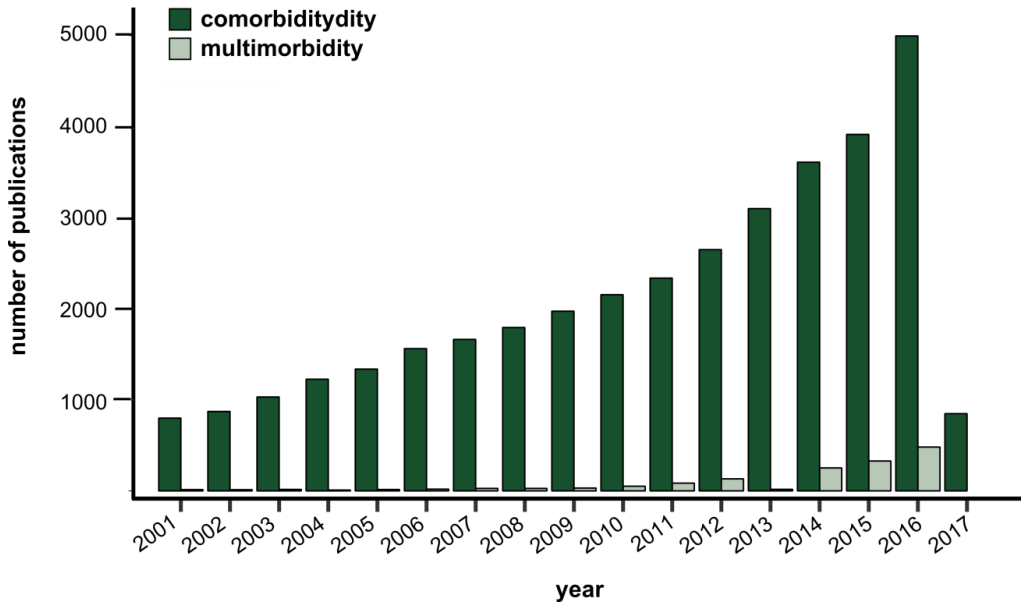


Figure 1 Comparative barplot between the number of publications in PubMed containing comorbidity or multimorbidity concept in the title of in the abstract, from 2001 to February 2017. The PubMed queries that retrieve this data are: comorbidity [Title/Abstract] and multimorbidity [Title/Abstract] respectively

The second most frequently term regarding the disease coexistence context in the literature is 'multimorbidity' (Figure 1). There is also a lack of consensus about its definition, and different attempts have tried also to unify them without success. In 2014, the International Research Community of Multimorbidity, in which they called to the research participation, trying to answer the question, "*which definition do you think*

should be used for multimorbidity?", being *"multiple co-occurring chronic or long term diseases or conditions, none considered as index disease¨*, the preferred multimorbidity definition (69%) between research participants all over the world.

To summarize, "comorbidity" is more often used when referring to the presence of multiple diseases in one individual, while "multimorbidity" is more often used in the context of multiple coexistence diseases without designation of an index one. The concept and the definition to choose for a specific study depend on the specific objectives of the study.

1.2 Comorbidity impact in population

In the era of big data and precision medicine, studies related to comorbidity have increased in an exponential way. Analyzing diseases co-occurring enhance the understanding of disease etiology (Tabarés-Seisdedos & Rubenstein 2013; Ibáñez et al. 2014).

1.2.1 Why is it important to study comorbidity?

Comorbidity potentially has implications for research, prevention and treatment, having a great importance for both, research and clinicians (Feinstein 1970; Jakovljević 2009; de Groot et al. 2003; Jakovljević & Ostojić 2013) and being an essential issue in personalized medicine.

1.2.1.1 Importance for research

Before we can begin to unravel the reasons behind any 'comorbidity', we need to carefully document the nature of any disease co-occurrence. This will give some insight into possible mechanisms underlying the association.

Comorbid disorders may have an important role in different types of research. The studies of comorbidity contribute to more complex knowledge about factors predisposing, promoting, establishing and maintaining disease in patients with several co-occurring disorders.

Comorbidity analysis can lead to discoveries new disease subtypes based on comorbidity patterns, being an impulse to research on the validity of current diagnostic systems (Starfield 2006).

1.2.1.2 Importance for treatment

Comorbidity is an extremely important issue in personalized medicine in regards to establishing more effective and efficient treatments (Jakovljević & Ostojić 2013).

If people who suffer from hypertension and asthma are more likely to have migraine, this needs to be taken in to account for determining the most appropriate treatment. For example, as suggested by Bonavita et al., while the best treatment for hypertensive patients with migraine are beta-blockers, in those patients affected by asthma they can precipitate in bronchoconstriction (Bonavita & De Simone 2008).

Therefore, characterizing patients according to their comorbidity patterns (Starfield 2006) would contribute to improve drug prescription, allowing having the most appropriate treatment outcome for each particular case, considering the patient as a whole entity and not treating each disorder separately.

1.2.1.3 Importance for prevention

Comorbidity has also significant implications for preventive medicine having two main broad implications: (i) an understanding of the nature of comorbidity will help dictate the targets of prevention, and (ii) if known, prevention efforts should be broad in the comorbidity target. If comorbidity arises because different problems or disorders share the same risk factors, then interventions addressing these risk factors should reduce the prevalence of these multiple problems.

Prevention programs have traditionally operated in isolation from each other. For example, it is often the case that programs addressing suicide and substance use prevention taking occur separately. There is rarely an attempt to conduct programs aimed at addressing multiple problems in an integrated fashion.

Understanding the nature of comorbidity will help better defining the targets of prevention. On-time interventions addressing shared risk factors should reduce the prevalence of various comorbidities. Only integrative and comprehensive prevention activities can achieve a satisfying success.

1.2.2 Comorbidity causes

Patients sharing molecular, environmental, and lifestyle risk factors may be at an increased risk of developing several other genetically and environmentally associated diseases in the same individual (Barabási et al. 2011; Davis et al. 2010). For instance, many well-known and influential lifestyle factors such as smoking, diet, and alcohol intake are actively related to diabetes type 1 and type 2, and obesity (Astrup 2001).

While the cause of comorbidities remains unknown in most of the cases, several hypotheses exist concerning the reasons why comorbidity might occur, including: (i) the causal relationship between them, (ii) common factors that could increase the risk of both diseases, (iii) common genetic origin and (iv) spurious relationships (Caron & Rutter 1991; Mueser et al. 1998). Moreover, recent evidence has exhibited that microRNAs and single nucleotide polymorphism (SNPs) also play key roles in the evolution and progression of human diseases (Jiang et al. 2010; Lewis et al. 2011).

1.2.3 Comorbidity prevalence

Community and clinical population studies show that comorbidity is a common phenomenon, the rule rather than the exception, particularly in the elderly, being a "normal state of affairs" for those aged over 65 years (Jakovljević & Ostojić 2013; Taylor et al. 2010). Although its prevalence rises with age, it is not a problem limited to the elderly population.

In Spain, the prevalence of comorbidity in patients older than 20 years old is 30 percent. Similar figures were reported for other European countries, with a higher prevalence of multimorbidities in people older than 50 years old.

According to some research, comorbidity is reported in 35 to 80% of all ill people (Taylor et al. 2010; Bonavita & De Simone 2008; Mezzich & Salloum 2008), and in elderly this fluctuate from 49% to 99% with an average number of chronic diseases per person between 2.5 and 6.5 (Fortin et al. 2005).

Is should be noted that the prevalence differs depending on: (i) the population under study (general population, hospitalized patients, the region, etc), (ii) the comorbidity definition, (iii) the comorbidity measurements selected.

1.3 Disorders and comorbidities

[Parrafo explicando cuales son las enfermedades que presentan mayor numero de comorbilidades, destacando que las psychiatrics son unas de las que mas comorbilidades presentan]

Comorbidity term has recently become very fashionable in psychiatry to indicate not only those cases in which a patient receives both a psychiatric and a general medical diagnosis (e.g. major depression and hypertension), but also those cases in which a patient receives two or more psychiatric diagnoses (e.g. major depression and panic disorder). This co-occurrence of two or more psychiatric diagnoses ('psychiatric comorbidity') has been reported to be very frequent.

For instance, in the US National Comorbidity Survey (Kessler et al, 1994), 51% of patients with a DSM–III–R/DSM–IV (American Psychiatric Association, 1987, 1994) diagnosis of major depression had at least one concomitant ('comorbid') anxiety disorder and only 26% of them had no concomitant ('comorbid') mental disorder. In a study based on data from the Australian National Survey of Mental Health and Well-Being (Andrews et al, 2002), 21% of people fulfilling DSM–IV criteria for any mental disorder met the criteria for three or more concomitant ('comorbid') disorders.(Zimmerman & Chelminski 2003)

This synthesis presents evidence that persons with comorbid mental and medical conditions represent just such a population. Based on epidemiological data from the 2001–2003 National Comorbidity Survey Replication, 34 million American adults, or 17 percent of the adult population, had comorbid mental and medical conditions within a 12-month period (3, 146). The high prevalence of this comorbidity, the complex causal connections linking medical and mental health conditions, and system fragmentation lead to problems in quality and costs related to comorbidity that are commonly even more complicated and burdensome than the problems related to the individual conditions themselves

Comorbidity between medical and mental conditions is the rule rather than the exception. In the 2001–2003 National Comorbidity Survey Replication (NCS-R), a

nationally representative epidemiological survey, more than 68 percent of adults with a mental disorder (diagnosed with a structured clinical interview) reported having at least one general medical disorder, and 29 percent of those with a medical disorder had a comorbid mental health condition

Studies examining the association between specific medical and mental disorders in nationally representative samples have found high rates of comorbidity. For example, in the 1996 Medical Expenditure Panel Survey, the risk of self-reported depression among people reporting diabetes was two times the risk for individuals without diabetes (50). In the 2006 Behavioral Risk Factor Survey, people reporting a diagnosis of asthma were 2.3 times more likely to screen positive for current depression compared with people without asthma (141). Conversely, in the 2001–2002 National Epidemiologic Survey on Alcohol and Related Conditions, persons reporting cardiovascular disease were at 1.43 times elevated risk of having a lifetime anxiety disorder (63).

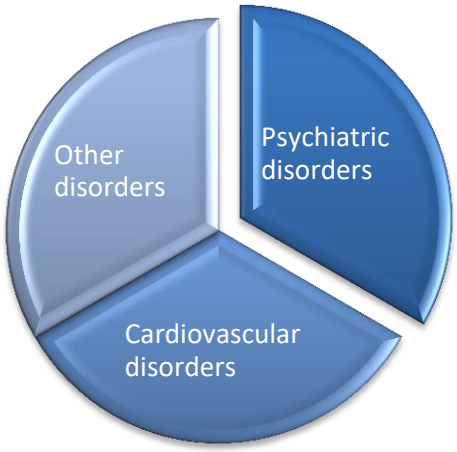
Medical disorders may lead to mental disorders, mental conditions may place a person at risk for medical disorders, and mental and medical disorders may share common risk factors. Epidemiological studies have been important in examining these pathways. For instance, medical conditions that are accompanied by a high symptom burden, such as migraine headaches or back pain, can lead to depression (116). At the same time, major depression is a risk factor for developing medical conditions, such as cardiovascular disease, that are characterized by pain or inflammation (118)

The average rheumatoid arthritis patient has approximately 1.6 comorbidities [104], and the number increases with the patient's age. As may be expected, the more comorbidities a patient has, the greater the utilization of health services, the greater societal and personal costs, the poorer the quality of life, and the greater chances of hospitalization and mortality. Moreover, comorbidity adds considerable complexity to patient care, making diagnosis and treatment decisions more challenging. For example, myocardial infarction (MI) is much more likely to be silent among persons with diabetes mellitus or RA, than in the absence of those comorbidities.

1.3.1 Psychiatric disorders

1.3.1.1. Depression

1.3.1.2. Substance use disorders
1.3.2. Comorbidity in psychiatry
1.3.2.1. Relevance of comorbidity of mental disorders in substance users
1.3.2.2. Prevalence and statistics about psychiatric comorbidities
1.3.2.3. Comorbidity between drug abuse and other mental illness
1.3.2.3. Mechanisms of the comorbidity of substance use and mental disorders



1.3.1 Psychiatric disorders

Psychiatric disorders have a great impact both on morbidity and mortality [1], [2]. According to the World Health Organization (WHO), depression, one of the most prevalent psychiatric disorders, affects 350 million people [3], [4]. Moreover, depression accounts for the biggest share of disease burden worldwide (76.4 million years lost to disability (YLD))[5].

Focusing on the case of comorbid mental and substance use disorders, different considerations must be taken into account. Comorbidity is the occurrence of more than one condition/disorder at the same time, and is common among those with mental illness. It can involve more than one mental disorder, or one mental disorder and one or more physical conditions. People with multiple disorders are more disabled and consume more health resources than those with only one disorder.

The relationship between mental and physical illness is not clearly understood, but having a physical illness is one of the strongest risk factors for depression (Wilhelm et al. 2003). Conversely, depression is also a risk factor for physical illness (Wulsin et al. 1999).

Understanding these complexities, and the best ways to treat each, requires consideration of both the mental and the physical illness (Clarke & Currie 2009).

1.3.1.1 Depressive disorders

Depressive disorders have a higher impact on the quality of life than most chronic physical conditions and are closely related to mortality [6], [7]. Projections indicate that after heart disease, depression is expected to become the second leading cause of disease burden by the year 2020, and the leading one in 2030 by the WHO [8].

The majority of individuals affected by depression will simultaneously meet diagnostic criteria for two or more additional health disorders [9]. In other words, depression patients usually suffer from comorbid diseases. Depression has been associated with a wide range of disorders, such as cardiovascular disease[14], diabetes[15], cancer[16], osteoporosis[17], arthritis[18], or substance use disorders [19]. Nevertheless, the nature and extent of the comorbidities of depression as well as their relation with age and gender at the population level need to be further studied [4, 15, 16].

It has been reported that many individuals with depression present comorbidities, being this condition the norm rather than the exception in depressive disorder patients, and they strongly depend on age and gender [4, 21, 23]

Depression is a debilitating disorder that disrupts relationships and daily lives and is one of the most common mental disorders(119). Depression is more common in women than in men (68, 119). People with depression may present as tearful and report that they feel sad, empty, hopeless and discouraged. Children and adolescents may present as irritable. Adults may also present as irritable and report concentration problems. People with depression may report loss of interest or pleasure in most activities, trouble sleeping, fatigue and problems with weight. Feelings of worthlessness and guilt may be associated with suicidal ideation.

To be diagnosed with major depression, a person must either have a depressed mood or a loss of interest or pleasure in daily activities consistently for at least a two week

period. This mood must represent a change from the person's normal mood. Social, occupational, educational or other important functioning must also be negatively impaired by the change in mood.

1.3.1.2 Substance use disorders

Various terms are used by different organizations (e.g. the EMCDDA, WHO) and in disease classifications to describe the more problematic forms of drug use. These terms include high-risk drug use, harmful use, substance abuse, substance dependence and, recently, in DSM-5 (APA, 2013), substance use disorder to different severity degrees.

1.3.2 Comorbidity between drug abuse and other mental illness

People with depression have high rates of comorbidity with other mental disorders and substance use disorders(17, 68, 119-122). Comorbidity in people with depression results in higher levels of impairment(119, 123) and increased severity and recurrence of depression(121, 122, 124). Depression and anxiety frequently co-exist(68, 119, 121, 122, 124).

Many people who regularly abuse drugs are also diagnosed with mental disorders and vice versa. The high prevalence of this comorbidity has been documented in multiple national population surveys since the 1980s. Data show that persons diagnosed with mood or anxiety disorders are about twice as likely to suffer also from a drug use disorder (abuse or dependence) compared with respondents in general. The same is true for those diagnosed with an antisocial syndrome, such as antisocial personality or conduct disorder. Similarly, persons diagnosed with drug disorders are roughly twice as likely to suffer also from mood and anxiety.

The high prevalence of comorbidity between drug use disorders and other mental illnesses does not mean that one caused the other, even if one appeared first. In fact, establishing causality or directionality is difficult for several reasons. Diagnosis of a mental disorder may not occur until symptoms have progressed to a specified level (per DSM); however, subclinical symptoms may also prompt drug use, and imperfect

recollections of when drug use or abuse started can create confusion as to which came first.

A particularly active area of comorbidity research involves the search for genes that might predispose individuals to develop both addiction and other mental illnesses, or to have a greater risk of a second disorder occurring after the first appears.

It is estimated that 40–60 percent of an individual's vulnerability to addiction is attributable to genetics; most of this vulnerability arises from complex interactions among multiple genes and from genetic interactions with environmental influences. In some instances, a gene product may act directly, as when a protein influences how a person responds to a drug (e.g., whether the drug experience is pleasurable or not) or how long a drug remains in the body. But genes can also act indirectly by altering how an individual responds to stress or by increasing the likelihood of risk-taking and novelty-seeking behaviors, which could influence the development of drug use disorders and other mental illnesses. Several regions of the human genome have been linked to increased risk of both drug use disorders and mental illness, including associations with greater vulnerability to adolescent drug dependence and conduct disorders. Furthermore, some areas of the brain are affected by both drug use disorders and other mental illnesses. For example, the circuits in the brain that use the neurotransmitter dopamine—a chemical that carries messages from one neuron to another— are typically affected by addictive substances and may also be involved in depression, schizophrenia, and other psychiatric disorders.

The largest study conducted so far has been carried was National Institute for Mental Health Epidemiological Catchment Area [ECA] Programme11. They carried out their study in a sample of 20, 291 subjects, used the Diagnostic Interviewing Schedule (DIS) and the DSM-III criteria. Among patients with alcohol dependence 36.6% had a comorbid psychiatric disorders out of which the common disorders were anxiety disorders (19.4%), antisocial personality disorders (14.3%), affective disorders (13.4%), and schizophrenia (3.8%). Among patients of opiod dependence, 65.2% had a comorbid psychiatric diagnosis; 31.6% anxiety disorders, 36.7% antisocial personality disorders, 30.8% affective disorders and 11.4% had schizophrenia.

Alcoholics are three times more likely to have another psychiatric disorder3. The self medication hypothesis for drug dependence proposed also signifies etiological relationship between the substance abuse and mental disorder4. The importance of this area can be recognized from the fact that nearly one hundred articles on this topic are being published in the indexed journals every year.

- ➔ Necesidad de extraer la informacion que se publica sobre enfermedades psychiatricas de la literatura y crear una base de datos como PsyGeNET

Within psychiatry, comorbidity is commonly used to refer to the overlap of two or more psychiatric disorders (Boyd, Burke, Gruenberg, et al., 1984). Comorbidity between substance use disorders and other mental disorders has gained increasing prominence in psychiatry and psychology within the past few decades (Wittchen, 1996).

Use of the term 'psychiatric comorbidity' to indicate the concomitance of two or more mental disorders might be incorrect because in most cases it is unclear whether the concomitant diagnoses actually reflect the presence of distinct clinical entities or refer to multiple manifestations of a single clinical entity. This multiplicity of psychiatric diagnoses may be explained, on the one hand, as a product of some specific features of current diagnostic systems for mental disorders. In this sense, psychiatric diagnoses are syndromes rather than diseases that have known physiopathology and valid and reliable biological markers (e.g. biochemical tests).

This lack of biological markers for psychiatric conditions has forced psychiatrists to develop operative diagnostic criteria, including the DSM (APA) and the ICD (WHO). Since the appearance of the DSM-III (APA, 1980) and subsequent DSM diagnostic criteria, mental disorders are diagnosed through a descriptive, categorical system that splits psychiatric behaviours and symptoms into numerous distinct diagnoses. Accordingly, the number of distinct psychiatric diagnoses described increased.

On the other hand, to increase the validity and reliability of psychiatric diagnoses, different diagnostic interviews have been designed to assess psychiatric disorders in a

systematic and standardised manner in accordance with main diagnostic criteria (such as those outline by the DSM and ICD) in order to eliminate biases. These interviews include the Schedule For Affective Disorders And Schizophrenia (SADS) (Endicott and Spitzer, 1978), the Structured Clinical Interview for DSM (SCID) (Spitzer et al., 1992) and the Composite International Diagnostic Interview (CIDI) (WHO, 1990). Their use reduces variability and improves diagnosis agreement and also helps to identify several clinical aspects that, in the past, tended to go unnoticed after the principal diagnosis had been made.

The high rate of comoRbidity between drug use disorders and other mental illnesses argues for a comprehensive approach to intervention that identifies and evaluates each disorder concurrently, providing treatment as needed. The needed approach calls for broad assessment tools that are less likely to result in a missed diagnosis. Accordingly, patients entering treatment for psychiatric illnesses should also be screened for substance use disorders and vice versa. Accurate diagnosis is complicated, however, by the similarities between drug-related symptoms such as withdrawal and those of potentially comorbid mental disorders. Thus, when people who abuse drugs enter treatment, it may be necessary to observe them after a period of abstinence in order to distinguish between the effects of substance intoxication or withdrawal and the symptoms of comorbid mental disorders.

1.3.3 Mechanisms of the comorbidity of substance use and mental disorders

Although convincing evidence supports a strong association between several mental disorders and substance use disorders, the nature of this relationship is complex and may vary depending on the particular mental disorder (e.g. depression, psychosis, post-traumatic stress disorder) and the substance in question (e.g. alcohol, cannabis, opioids, cocaine).

As mentioned previously, there are three main ways in which different diseases or disorders may occur in the same individual: chance, selection bias or causal association. Focusing on the comorbidity of substance use and mental disorders, we list below four nonexclusive aetiological and neurobiological hypotheses that could explain comorbidity.

The first hypothesis is that the combination of a substance use and another mental disorder may represent two or more independent conditions. In this case, the combination may occur through chance alone (roughly, the prevalence of one disorder multiplied by the prevalence of the other) or as a consequence of the same predisposing factors (e.g. stress, personality, childhood environment, genetic influences) that affect the risk for multiple conditions. That is, substance use disorders and other psychiatric disorders would represent different symptomatic expressions of similar pre-existing neurobiological abnormalities (Brady and Sinha, 2005).

Research in basic neuroscience has demonstrated the key roles of biological and genetic or epigenetic factors in an individual's vulnerability to these disorders. Genes, neural bases and environment are intimately interconnected. All psychoactive substances with abuse potential have a counterpart in, or correspond to, some endogenous system, such as the opioid system, the endocannabinoid system, the cholinergic/nicotínic system or the dopaminergic system. An inherited or acquired deficiency in these neurobiological Systems and circuits may explain addictive behaviour and other psychiatric symptoms.

The second hypothesis is that the psychiatric disorder other than the substance use disorder is a risk factor for drug use and the development of a comorbid substance use disorder. In this scenario, different situations can be considered. In the 'self-medication hypothesis' (Khantzian, 1985), the substance use disorder develops as a result of attempts by the patient to deal with problems associated with the mental disorder (e.g. social phobia, post-traumatic stress disorder, psychosis). In this case, the substance use disorder might become a long-term problem, or the excessive use of alcohol or an illicit drug.

(...)

1.3.4 Relevance of comorbidity of mental disorders in substance users

Apart from the difficulties in defining and diagnosing psychiatric comorbidity in those individuals with a substance use disorder, another crucial aspect is their impact, not only on clinical care but also on Health service planning and financing.

Dual diagnosis has been associated with poor outcomes in affected subjects. In comparison with patients with a single disorder, dually diagnosed patients show a

higher psychopathological severity, more emergency admissions (Booth et al., 2011; Curran et al., 2008; Langås et al., 2011; Martín-Santos et al., 2006; Schmoll et al., 2015), significantly increased rates of psychiatric hospitalisation (Lambert et al., 2003; Stahler et al., 2009) and a higher prevalence of suicide (Aharonovich et al., 2006; Conner, 2011; Marmorstein, 2011; Nordentoft et al., 2011; Szerman et al., 2012).

In addition, comorbid drug users show increased rates of risky behaviours, which are linked to infections, such as HIV (human immunodeficiency virus) and hepatitis B and C viruses (Carey et al., 2001; Durvasula and Miller, 2014; Khalsa et al., 2008; King et al., 2000; Loftis et al., 2006; Rosenberg et al., 2001), as well as psychosocial impairments, such as higher unemployment and homelessness rates (Caton et al., 1994; Krausz et al., 2013; Vázquez et al., 1997), and considerable violent or criminal behaviour (Abram and Teplin, 1991; Cuffel et al., 1994; Greenberg and Rosenheck, 2014; Soyka, 2000).

Taking into account the burden on health and legal systems, psychiatric comorbidity among drug users leads to high costs for society (DeLorenze et al., 2014; Whiteford et al., 2013). Clinical practice research has shown that comorbid disorders are reciprocally interactive and cyclical, and poor prognoses for both psychiatric and substance use disorders can be expected if treatment does not tackle both (Boden and Moos, 2009; Flynn and Brown, 2008; Magura et al., 2009). Treatment of dual diagnosis patients is set to be one of the biggest challenges in the drugs field in the coming years. The key questions will revolve around where, how and for how long to treat these patients (Torrens et al., 2012).

1.4 Electronic Health Records

Electronic health record (EHR) systems enable hospitals to store and retrieve detailed patient information to be used by health care providers, and sometimes patients, during a patient's hospitalization, over time, and across care settings. Embedded clinical decision support and other tools have the potential to help clinicians provide safer, more effective care than is possible by relying on memory and paper-based systems. In addition, EHRs can help hospitals monitor, improve, and report data on health care quality and safety.

Many challenges exist when it comes to repurposing data from an electronic medical record system (EMRS) for research. (...) The repurposing of medical record data for clinical research holds high promise.

The growth of electronic health records (EHRs) has been recognized as a viable and efficient model for genetic research. EHRs represent an unprecedented opportunity to leverage clinical data generated as a byproduct of healthcare for genetic discovery. (*Extracting research-quality phenotypes from electronic health records to support precision medicine*)

HER system is primarily designed for routine clinical care. EHRs contain a wealth of clinical information, but this information is not always in readily minable formats. Designed for clinical care, diagnoses may only be mentioned in clinical notes. Thus, to identify populations with high accuracy takes careful thought and domain knowledge. (*Extracting research-quality phenotypes from electronic health records to support precision medicine*)

Domain experts understand the target phenotype and its representation, while clinical informaticians know where and how to extract corresponding information. Validation is another important part of the process that not only measures an algorithm performance, but also enhances its capability for inter-institutional sharing. (*Extracting research-quality phenotypes from electronic health records to support precision medicine*)

EHR data come in both structured and unstructured formats, and the use of both types of information can be essential for creating accurate phenotypes.

Figure showing the HER data typical structure

1.4.1 EHRs advantages

EHRs have several distinct advantages for genetic research, including:

- cost efficiency
- the large amounts of available clinical data
- the ability to analyze data over time
- eliminate the cost of recruiting patients for each phenotype study
- HER-based studies took much shorter time than traditional research designs to complete
- Availability of longitudinal clinical information is an asset for genetic reseach

1.5 Comorbidity measurements

Before studying determinants or consequences of multiple pathology a number of conceptual decisions have to be made.

a) Comorbidity definition

The first decision to be made is to choose between focusing on a specific disease with accompanying conditions or on co-occurring diseases in general, without any hierarchical order. In other words, studying comorbidity—that is "the existence or occurrence of any distinct additional entity during the clinical course of a patient who has the index disease under study" — or studying multimorbidity, "the cooccurrence of multiple chronic or acute diseases and medical conditions within one person". (Akker, 2000)

b) Set of diseases analyzed

The set of diseases included in the operationalization and analysis of comorbidity and multimorbidity is decisive for the amount of comorbidity or multimorbidity found.

Kolnaar et al. Depending on the research subject, in comorbidity research it might be useful to include chronic as well as acute problems. Furthermore, it should be noted that applying unusual selection criteria for disorders hampers comparison of the results to other studies. (Akker, 2000)

Generally, the higher the number of diseases included in the calculation, the higher the frequency of occurrence of comorbidity or multimorbidity. For example, the proportion of subjects in the XXXX database aged XXX or older suffering from two or more disorders from a list of four (hypertension, emphysema, psoriasis and osteoporosis) was XXX%. When glaucoma, diabetes and gout were added to this selection of diseases, the prevalence of multimorbidity in the same population increased to XXXX%.

c) Population under study
Additionally, the diagnostic criteria and diagnostic procedures that are used are strongly related to the number of diseases identified. For example, a disease will be identified more frequently if the diagnostic criteria are less strict or when active case-finding is used. This stresses the importance of proper description of definitions, procedures and criteria used to facilitate the correct understanding of and comparison among different studies.

Restrictions regarding age groups or special care settings have a large impact on study results. Differences illustrate the considerable impact of choices made regarding the inclusion and exclusion of diseases in analysis. It once more shows the importance of an adequate description of diseases that are selected for a specific study.

1.5.1 Comorbidity measurements

Fisher test A Fisher exact test for each pair of diseases is performed to assess the null hypothesis of independence between the two diseases. Four groups of patients are defined in order to perform the statistical testing: patients suffering disease A and disease B, patients suffering disease A but not disease B, patients suffering disease B but not disease A and patients not suffering disease A nor disease B. The Fisher exact test is then applied to estimate the p-value for each pair of diseases. The Benjamini-

Hochberg false discovery rate method [13] is applied on the ranked list to correct for multiple testing.

Comorbidity score

$$comorbidityscore = log_2\left(\frac{observed+1}{expected+1}\right) expected = \frac{n_A n_B}{N} \quad (1)$$

where observed stands for the number of disease-disease associations (disease A and disease B), and expected is estimated based on the occurrence of each disease (number of patients diagnosed with disease A, nA, multiplied by the number of patients diagnosed with the comorbid disorder B, nB , and divided by the total number of patients, N). Since logarithm is applied, a comorbidity score of 1.0 means that the observed comorbidities are higher than two fold (approximately) than expected.

Relative Risk (RR): The relative risk (RR) expresses the relationship between the rate of incidence of a disease among the patients exposed and those patients that are not exposed to a certain risk factor. Here the risk factor is other disease. The RR value moves from 0 to infinite.

- RR = 1: The risk is the same for patients exposed to the factor of risk than for those patients that are not exposed.
- RR > 1: The patients exposed to the factor of risk are more likely to su_er the disease. It is a risk factor.
- RR < 1: The patients exposed to the factor of risk are less likely to su_er the disease. It is a protection factor.

The RR is estimated as the fraction between the number of patients diagnosed with both diseases and random expectation based on disease prevalence. The RR of observing a pair of diseases A and B affecting the same patient is given by [2]:

$$RR_{ij} = \frac{C_{ij}N}{P_i P_j}$$

where Cij is the number of patients affected by both diseases, N is the total number of patients in the population and Pi and Pj are the prevalences of diseases i and j.

ϕ-correlation: the Pearsons correlation for binary variables measures the robustness of the comorbidity association (Moni and Lio, 2014). The ϕ-correlation, which is Pearson's correlation for binary variables, can be expressed mathematically as:

$$\phi_{ij} = \frac{C_{ij}N - P_i P_j}{\sqrt{P_i P_j (N - P_i)(N - P_j)}}$$

where N is the total number of patients in the population, PA i and PB are incidences/prevalences of diseases A and B respectively. CAB is the number of patients that have been diagnosed with both diseases A and B, and PA PB is the random expectation based on disease prevalence. The Pearson correlation coefficient, can take a range of values from +1 to -1. A value of 0 indicates that there is no correlation between the two diseases; a value greater than 0 indicates a positive correlation between the two diseases and a value less than 0 indicates a negative correlation.

Odds ratio The odds ratio represents the increased chance that someone suffering disease A will have the comorbid disorder B. It shows the extent to which suffering a disorder increases the risk of developing another illness or disorder. The odds ratio is derived from a comparison of rates of the illness among individuals who do and do not exhibit the factor of interest. A statistically significant odds ratio (significantly different from 1.00 at the .05 level) indicates an appreciable risk associated with a particular factor. For example, an odds ratio of 2.00 indicates a doubled risk of the appearance of the disorder.

These measures allow us to quantify the co-occurrence of disease pairs compared with the random expectation. The user can select the measure and the cut-off value in order to assess disease comorbidity

The comorbidity measures are not completely independent of each other. For example, RR and phi, they both increase with the number of patients affected by both diseases, yet both measures have their intrinsic biases. For example, RR overestimates relationships involving rare diseases and underestimates the comorbidity between highly prevalent illnesses, whereas w accurately discriminates comorbidities between

pairs of diseases of similar prevalence but underestimates the comorbidity between rare and common diseases.

When two diseases co-occur more frequently than expected by chance, we will get $RR_{ij} > 1$ and $\phi_{ij} > 0$.

1.6 Comobidity tools/softwares

Exploringdisease-disease associations by using multi-omics and clinical information is expected to improve our current knowledge of disease relationships, which may lead to further improvements in disease diagnosis, prognosis and treatment (Park et al., 2009). Recent research has increasingly demonstrated that many seemingly dissimilar diseases have common molecular mechanisms and strong associations among them (Yu andWang, 2015).

1.6.1 POGO R software (Ali Moni & Pietro Lio, 2015)

POGO computes statistically significant associations among diseases, to predict disease comorbidity risk and to develop comorbidity maps, which are useful for the physicians and informative for the patients. To perform the computation of the comorbidity risk, this software uses clinical, gene expression, miRNA, SNPs, CNVs, ontology, phenotypic, and environmental data. The inputs of this software is the initial diagnostic result of the patient.

The goal of this software is to construct comorbidity maps that incorporate disease interactions, omics, phenotypic and ontology information, and environmental influences. It is a user-friendly and interactive personalised disease and disease comorbidity prediction software. It provides different comorbidity assessment and stratification; integration of omics information with POGO output data could be used to predict more accurate survival probability of patients. Clinical information is from

the http://www.icd9data.com in the ICD-9-CM format and collected from Hidalgo et al. (2009).

Two diseases are connected if they are co-expressed in a significant number of patients in a population (Hidalgo et al., 2009). To estimate the correlation starting from disease co-occurrence, we need to quantify the strength of the comorbidity risk. We used two comorbidity measures to quantify the strength of comorbidity associations between two diseases: (i) the Relative Risk and (ii) ϕ-correlation (Pearsons correlation for binary variables).

1.7 Comorbidity research examples

1.7.1 Comorbidity network

We can summarize the set of all comorbidity associations between all diseases expressed in the study population by constructing a Phenotypic Disease Network (PDN). In the PDN, nodes are disease phenotypes identified by unique ICD9 codes, and links connect phenotypes that show significant comorbidity according to the measures introduced above (Hidalgo, 2009).

A comorbidity relation means that two diseases occur more frequently within a patient than what would be expected from the frequency of the individual diseases alone. This means that the joint probability for suffering two diseases i and j together is larger than the product of the probabilities of the individual diseases (prevalences). (Chmiel, 2014)

$$P(i, j) > P(i)P(j)$$

A phenotypic human disease network (PHDN) consists of nodes representing the diseases and links that indicate comorbidity relations (Chmiel, 2014).

1.8 Comorbidity explanation

Note that only a limited number of comorbidities can be explained by common genes, proteins, or metabolites. These differences with respect to the clinical reality of disease not only reflect our limited knowledge of cellular processes, they also underscore the role of environmental and epigenetic factors in disease progression (Chmiel, 2014).

2 Objectives

“If there is a good will, there is great way”

William Shakespeare (1564 - 1616)

Biomedical research on the big data era can enable to predict the risk of suffering one disorder based in the other disorder you have, and give to the patient the best treatment for the bunch of disorders he suffer, walking through the personalized medicine.

In order to harness the power of emerging disease data, the aim of this thesis is to develop a system to collect and transform this information in knowledge that could be analyzed, interpreted by doctors, and applied in order to improve patients' life. In particular, the specific goals are:

(1) To develop a methodology for the exploitation of clinical data from patient record databases to identify comorbid diseases and to detect comorbidity patterns.

(2) To develop a gene-disease association database for psychiatric disorders to analyze comorbidities from the genomic point of view, exploring the molecular and cellular mechanism that substantiate psychiatric disorder comorbidities.

(3) To apply the software developed to different databases and local population registers, to analyze comorbidity patterns and their correlation with the molecular data.

(4) Apply the developed methodology to the comorbidity study of disorders like Chronic obstructive pulmonary disease (COPD), depressive disorder or substance drug induced disorders in order to:
    a) Determine the comorbidity prevalence in the Catalan population
    b) Analyze the biological mechanisms that substantiate these comorbidities.

3 The comorbidity R package

4 Applications and results

4.1 PsyGeNET: a knowledge platform on psychiatric disorders and their genes.

Comorbidity is the norm among common mental disorders, as more than 50% of affected people meet criteria for multiple diseases. The coexistence of mood and substance use disorders (SUD) is attracting growing interest in the scientific community because of its high prevalence rates and its association with a greater severity of illness and rate of recurrence for both disorders. In particular, alcohol and cocaine dependences are frequently associated to depression. Several mechanisms have been proposed to explain the coexistence of diseases in one patient, being the genetic origin one of them. With the objective of having a curated gene-disease resource focus on psychiatric disorders, PsyGeNET was developed. PsyGeNET is a new resource that integrates information on psychiatric disorders and their genes, offering exploratory tools for the analysis of gene-disease associations. Due to its special focus on psychiatric diseases, comprehensiveness and high-quality database, PsyGeNET represents a valuable resource for the analysis of the molecular underpinning of psychiatric disorders and their comorbidities

Gutiérrez-Sacristán, A., Grosdidier, S., Valverde, O., Torrens, M., Bravo, À., Piñero, J, F. Sanz & Furlong, L. I. (2015). *PsyGeNET: a knowledge platform on psychiatric disorders and their genes*. Bioinformatics, 2015 May, Vol: 31 (18) pp: 3075-

4.2 Molecular and clinical diseasome of comorbidities in exacerbated COPD patients

The frequent occurrence of comorbidities in patients with chronic obstructive pulmonary disease (COPD) suggests that they may share pathobiological processes and/or risk factors. To explore these possibilities we compared the clinical diseasome and the molecular diseasome of 5447 COPD patients hospitalized because of an exacerbation of the disease. The clinical diseasome is a network representation of the relationships between diseases, in which diseases are connected if they co-occur more than expected at random; in the molecular diseasome, diseases are linked if they share associated genes or interaction between proteins. The results showed that about half of the disease pairs identified in the clinical diseasome had a biological counterpart in the molecular diseasome, particularly those related to inflammation and vascular tone regulation. Interestingly, the clinical diseasome of these patients appears independent of age, cumulative smoking exposure or severity of airflow limitation. These results support the existence of shared molecular mechanisms among comorbidities in COPD.

Faner, R., Gutiérrez-Sacristán, A., Castro-Acosta, A., Grosdidier, S., Gan, W., Sánchez-Mayor, M., Lopez-Campos, JL., Pozo-Rodriguez, F., Sanz, F., Manino, D., Furlong, L. I. & Agusti A. (2015). *Molecular and clinical diseasome of comorbidities in exacerbated COPD patients*. European Respiratory Journal, ERJ- 2015, June Vol: 46 (4)

5 Discussion

"If we knew what it was we were doing, it would not be called research, would it?"
Albert Einstein (1879-1955)

5.1 Overview

Multiple pathology is a health problem that is increasingly important in research because of its prevalence and impact. Possible options, both in the design and in the analysis of studies on comorbidity and multimorbidity, should be well considered in advance (Marjan van den Akker, 2000)

5.1.1 Comorbidity definition

There are many terms currently used in the context of múltiple concurrent diseases. This diversity may have a negative impact on practice and research. The term comorbidity should be reserved for situations in which one or more diseases coexist with an index disease under study. For situations in which there are multiple coexistent diseases, but none considered as index disease, the term multimorbidity was by far the most frequently used. The term multimorbidity was most often defined as “the presence of more than one or multiple chronic or long-term diseases or conditions”. We recommend to clearly definí terms until a general consensus on terminology of múltiple coexistent diseases is reached (Almirall, 2013).

General agreement about terminology and definitions should be reached. We suggest that only two terms be used, one for situations in which multiple diseases/conditions coexist with an index disease under study, and another for situations in which there are multiple coexistent diseases/conditions, but none considered as the index disease/condition. In the first case, the term comorbidity already exists and is widely accepted and used. Indeed, at the time of writing this text, it is the only term, of those mentioned above, that is accepted as a Medical Subject Heading (MeSH) (Tree numbers: N05.715.350.225; N06.850.490.687) by the US National Library of Medicine (http://www.nlm. nih.gov/mesh/MBrowser.html), and is defined as “the presence of co-existing or additional diseases with reference to an initial diagnosis or with reference to the index condition that is the subject of study…”.

Redefining or using the term comorbidity for other purposes, particularly in situations where no index disease is considered, should be avoided. Publications in which comorbidity is used with a different meaning from the one defined by Feinstein and adopted

by the US National Library of Medicine would be very difficult to find among thousands of other articles using the term as originally defined.

This use of the term ‘comorbidity’ to indicate the concomitance of two or more psychiatric diagnoses appears incorrect because in most cases it is unclear whether the concomitant diagnoses actually reflect the presence of distinct clinical entities or refer to multiple manifestations of a single clinical entity. Because ‘the use of imprecise language may lead to correspondingly imprecise thinking’ (Lilienfeld et al, 1994), this usage of the term ‘comorbidity’ should probably be avoided.

5.1.2 Disease selection for comorbidity studies

Obviously, the use of highly prevalent diseases has statistical advantages, since co-occurrences of diseases with low prevalences are rare .

5.1.3 Confounding variables when analyzing comorbidities

An important point of concern when analyzing comorbidity is the influence of effect modifying or confounding variables. For example, as age is a strong determinant of many diseases, it is generally important to take this variable into consideration when analyzing the co-occurrence of diseases. Part of the co-occurrence of diseases can be explained by known influences of age (e.g., benign prostate hypertrophy and osteoarthritis). Of course other variables such as socioeconomic status, environmental factors and psychological features can also be very influential. If these influences are not taken into account or at least described, this can lead to unrealistic or irrelevant outcomes (Marjan van den Akker, 2000).

An important conceptual consideration in this context is whether the covariable is an element of the causal chain to be evaluated [when (over) adjustment is not appropriate] or just a confounder without relevance to the causal chain of primary interest (when adjustment is useful). Evaluation of effect modification may be helpful to identify different co-occurrence patterns in various subgroups (Marjan van den Akker, 2000).
When analyzing combinations of two diseases, known covariables can be adjusted by using a multiple regression analysis (using one of the diseases as the dependent variable) or a stratified analysis according to Mantel–Haenszel. When analyzing

combinations of three or more cooccurring diseases, stratified analyses are a good option as long as the study population is sufficiently large, giving the opportunity to account for the main covariables. Another option is to carry out a stepwise multiple logistic regression analysis, evaluating the determinants of the presence of a disease additional to a specific disease or combination of diseases.

5.2 Perspectives for Medical informatics

The use of clinical information systems has substantially increased within the last decade, leading to a widespread introduction of electronic management systems for medical records. Nevertheless, to many clinicians those systems still appear more as a burden than as a support tool for their daily work.

5.2.1 Challenges: Linking Electronic Medical Records with Clinical Research Databases

Since inefficiencies in clinical trial data collection cause delays, increase costs, and may reduce clinician participation in medical research, electronic medical records (EMRs) have often been cited as a significant new tool for advancing clinical trial capabilities into standard clinical practice. However, we need to be aware of the fact that combining clinical research and clinical care activities into one unified electronic information System requires integrating a substantial body of regulatory requirements and institutional policies.

5.3 Future Perspectives

5.3.1 Genotype and phenotype data: a step further

We are only focus our attention in co-occurrence. Following the biology dogma, the ideal will be, in a near future, having the DNA data from the patients, and being able to go from genotype to phenotype and determine not only co-occurrence, what is more, causality. Although it was in 2000 when the code of human genome was cracked, we are in the very beginning of learning how to deal with these huge amount of data, but connecting the genotypic data with the diagnosis and phenotypic one, will be a great step towards the personalize medicine.

6 Conclusions

Now this is not the end.
It is not even the beginning of the end.
But it is, perhaps, the end of the beginning
Sir Winston Churchill (1874 - 1965)

The main achievements of this thesis are presented below.

(1) The `comoRbidity` R package approach was developed as a tool to extract and analyze comorbidity patterns from electronic health records.

7 Appendix

8 Bibliography

Van Den Akker, M. et al., 2001. Problems in determining occurrence rates of multimorbidity. *Journal of Clinical Epidemiology*, 54(7), pp.675–679.

van den Akker, M., Buntinx, F. & Knottnerus, J.A., 1996. Comorbidity or multimorbidity: what's in a name? A review of literature. *European Journal of General Practice*, 2, pp.65–70.

Almirall, J., Fortin, M. & Fortin, M., 2013. The coexistence of terms to describe the presence of multiple concurrent diseases. *Journal of Comorbidity*, 3(1), pp.4–9. Available at: http://www.jcomorbidity.com/index.php/test/article/view/22 [Accessed October 16, 2016].

Astrup, A., 2001. Healthy lifestyles in Europe: prevention of obesity and type II diabetes by diet and physical activity. *Public health nutrition*, 4(2B), pp.499–515. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11683545 [Accessed February 20, 2017].

Barabási, A.-L., Gulbahce, N. & Loscalzo, J., 2011. Network medicine: a network-based approach to human disease. *Nature reviews. Genetics*, 12(1), pp.56–68. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3140052&tool=pmcentrez&rendertype=abstract [Accessed July 11, 2014].

Bayliss, E.A. et al., 2008. Processes of care desired by elderly patients with multimorbidities. *Family Practice*, 25(4), pp.287–293. Available at: https://academic.oup.com/fampra/article-lookup/doi/10.1093/fampra/cmn040 [Accessed February 17, 2017].

Bonavita, V. & De Simone, R., 2008. Towards a definition of comorbidity in the light of clinical complexity. *Neurological Sciences*, 29(S1), pp.99–102. Available at: http://link.springer.com/10.1007/s10072-008-0898-1 [Accessed November 17, 2016].

Caron, C. & Rutter, M., 1991. Comorbidity in child psychopathology: concepts, issues and research strategies. *Journal of child psychology and psychiatry, and allied disciplines*, 32(7), pp.1063–80. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1787137 [Accessed February 20, 2017].

Cesari, M. et al., 2006. Comorbidity and Physical Function: Results from the Aging and Longevity Study in the Sirente Geographic Area (ilSIRENTE Study). *Gerontology*, 52(1), pp.24–32. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16439821 [Accessed February 17, 2017].

Collins, F.S. & Varmus, H., 2015. A New Initiative on Precision Medicine. *New England Journal of Medicine*, 372(9), pp.793–5.

Davis, D.A. et al., 2010. Time to CARE: a collaborative engine for practical disease prediction. *Data Mining and Knowledge Discovery*, 20(3), pp.388–415. Available at: http://link.springer.com/10.1007/s10618-009-0156-z [Accessed November 17, 2016].

Feinstein, A.R., 1970. The pre-therapeutic classification of co-morbidity in chronic disease. *Journal of Chronic Diseases*, 23(7), pp.455–468. Available at: http://www.sciencedirect.com/science/article/pii/0021968170900548 [Accessed February 12, 2015].

Fortin, M. et al., 2005. Multimorbidity is common to family practice: is it commonly researched? *Canadian family physician Medecin de famille canadien*, 51, pp.244–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16926936 [Accessed November 17, 2016].

Frank, L., 2000. When an Entire Country Is a Cohort. *Science*, 287(5462), pp.2398–2399.

Goh, K.-I. et al., 2007. The human disease network. *Proceedings of the National Academy of Sciences of the United States of America*, 104(21), pp.8685–90. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1885563&tool=pmcentrez&rendertype=abstract [Accessed January 15, 2015].

de Groot, V. et al., 2003. How to measure comorbidity. a critical review of available methods. *Journal of clinical epidemiology*, 56(3), pp.221–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12725876 [Accessed February 13, 2017].

Hamosh, A. et al., 2005. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic acids research*, 33(Database issue), pp.D514-7.

Hidalgo, C.A. et al., 2009. A dynamic network approach for the study of human phenotypes. *PLoS computational biology*, 5(4), p.e1000353. Available at: http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1000353 [Accessed May 25, 2015].

Ibáñez, K. et al., 2014. Molecular Evidence for the Inverse Comorbidity between Central Nervous System Disorders and Cancers Detected by Transcriptomic Meta-analyses. *PLoS Genetics*, 10(2).

Jakovljević, M., 2009. Psychopharmacotherapy and comorbidity: Conceptual and epistemiological issues, dilemmas and controversies. *Psychiatria Danubina*, 21(3), pp.333–340.

Jakovljević, M. & Ostojić, L., 2013. Comorbidity and multimorbidity in medicine today: challenges and opportunities for bringing separated branches of medicine closer to each other. *Psychiatria Danubina*, 25 Suppl 1(1), pp.18–28.

Jiang, Q. et al., 2010. Prioritization of disease microRNAs through a human phenome-microRNAome network. *BMC Systems Biology*, 4(Suppl 1), p.S2. Available at: http://bmcsystbiol.biomedcentral.com/articles/10.1186/1752-0509-4-S1-S2 [Accessed February 20, 2017].

Lawson, K.D. et al., 2013. Double trouble: the impact of multimorbidity and deprivation on preference-weighted health related quality of life a cross sectional analysis of the Scottish Health Survey. *International Journal for Equity in Health*, 12(1), p.67. Available at:

http://equityhealthj.biomedcentral.com/articles/10.1186/1475-9276-12-67 [Accessed February 17, 2017].

Lewis, S.N. et al., 2011. Prediction of disease and phenotype associations from genome-wide association studies. *PloS one*, 6(11), p.e27175. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22076134 [Accessed February 20, 2017].

Lynge, E., Sandegaard, J.L. & Rebolj, M., 2011. The Danish National Patient Register. *Scandinavian journal of public health*, 39(7 Suppl), pp.30–3.

Marengoni, A. et al., 2008. Prevalence of Chronic Diseases and Multimorbidity Among the Elderly Population in Sweden. *American Journal of Public Health*, 98(7), pp.1198–1200. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18511722 [Accessed February 17, 2017].

McGee, D. et al., 1996. Patterns of comorbidity and mortality risk in blacks and whites. *Annals of epidemiology*, 6(5), pp.381–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8915468 [Accessed February 17, 2017].

Mezzich, J.E. & Salloum, I.M., 2008. Clinical complexity and person-centered integrative diagnosis. *World psychiatry : official journal of the World Psychiatric Association (WPA)*, 7(1), pp.1–2. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18458769 [Accessed February 20, 2017].

Mikuls, T.R. & Saag, K.G., 2001. COMORBIDITY IN RHEUMATOID ARTHRITIS. *Rheumatic Disease Clinics of North America*, 27(2), pp.283–303.

Mueser, K.T., Drake, R.E. & Wallach, M.A., 1998. Dual diagnosis: A review of etiological theories. *Addictive Behaviors*, 23(6), pp.717–734.

Park, J. et al., 2009. The impact of cellular networks on disease comorbidity. *Molecular systems biology*, 5, p.262.

Pletscher-Frankild, S. et al., 2015. DISEASES: text mining and data integration of disease-gene associations. *Methods (San Diego, Calif.)*, 74, pp.83–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25484339 [Accessed October 4, 2016].

Schellevis, F.G. et al., 1993. Comorbidity of chronic diseases in general practice. *Journal of clinical epidemiology*, 46(5), pp.469–73. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8501473 [Accessed February 17, 2017].

Starfield, B., 2006. Threads and yarns: weaving the tapestry of comorbidity. *Annals of family medicine*, 4(2), pp.101–3. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16569711 [Accessed February 20, 2017].

Starfield, B. & Kinder, K., 2011. Multimorbidity and its measurement. *Health policy (Amsterdam, Netherlands)*, 103(1), pp.3–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21963153 [Accessed October 16, 2016].

Tabarés-Seisdedos, R. & Rubenstein, J.L., 2013. Inverse cancer comorbidity: a serendipitous opportunity to gain insight into CNS disorders. *Nature reviews. Neuroscience*, 14(4), pp.293–304.

Taylor, A.W. et al., 2010. Multimorbidity - not just an older person's issue. Results

from an Australian biomedical study. *BMC Public Health*, 10(1), p.718. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21092218 [Accessed February 20, 2017].

Valderas, J.M. et al., 2009. Defining comorbidity: implications for understanding health and health services. *Annals of family medicine*, 7(4), pp.357–63. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2713155&tool=pmcentrez&rendertype=abstract [Accessed January 22, 2015].

Verbrugge, L.M., Lepkowski, J.M. & Imanaka, Y., 1989. Comorbidity and its impact on disability. *The Milbank quarterly*, 67(3–4), pp.450–84. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2534562 [Accessed February 17, 2017].

Zimmerman, M. & Chelminski, I., 2003. Generalized anxiety disorder in patients with major depression: is DSM-IV's hierarchy correct? *The American journal of psychiatry*, 160(3), pp.504–12. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12611832 [Accessed March 8, 2017].

9 Appendix: Glossary

Addiction: A chronic, relapsing disease characterized by compulsive drug seeking and use and by long-lasting changes in the brain.

Bipolar Disorder: A mood disorder characterized by alternating episodes of depression and mania or hypomania.

Chronic disease: any disease which has one or more of the following characteristics: they are permanent, leave residual disability; they are caused by non-reversible pathological alteration, require special training of the patient for rehabilitation, or may be expected to require a long period of supervision, observation or care.

Comorbidity: The occurrence of two disorders or illnesses in the same person, either at the same time (co-occurring comorbid conditions) or with a time difference between the initial occurrence of one and the initial occurrence of the other (sequentially comorbid conditions).

Depression: A disorder marked by sadness, inactivity, difficulty with thinking and concentration, significant increase or decrease in appetite and time spent sleeping, feelings of dejection and hopelessness, and, sometimes, suicidal thoughts or an attempt to commit suicide.

Disorder: An abnormal condition of an organism which interrupts the normal bodily functions that often leads to feeling of pain and weakness, and usually associated with symptoms and signs. // A pathologic condition in which the normal functioning of an organism or body is impaired or disrupted resulting in extreme pain, dysfunction, distress, or death. // A disturbance in physical or mental health or functions. // A definite pathologic process with a characteristic set of signs and symptoms

Disease is defined as a pathologic process with a characteristic set of signs and symptons. It may affect the whole body or any of its parts, and its etiology, pathology

and prognosis may be known or unknown (MeSH). // Any condition in an organism that is other tha the healthy state.

Health: the World Health Organization's claims that health is a "state of complete physical, mental and social well-being, not merely the absence of disease or infirmity" (WHO, 1946)

Major Depressive Disorder: A mood disorder having a clinical course of one or more serious depression episodes that last 2 or more weeks. Episodes are characterized by a loss of interest or pleasure in almost all activities; disturbances in appetite, sleep, or psychomotor functioning; a decrease in energy; difficulties in thinking or making decisions; loss of self-esteem or feelings of guilt; and suicidal thoughts or attempts.

Mental Disorder: A mental condition marked primarily by sufficient disorganization of personality, mind, and emotions to seriously impair the normal psychological or behavioral functioning of the individual. Addiction is a mental disorder.

Medical Subject Headings (MeSH): terminology by which the world Medical literature is indexed. MeSH arranges terms in a structure that breaks from strict hierarchy used by most other coding schemes. Terms are organized intro hierchies and may appear in multiple places in the hierarchy.

Multimorbidity: co-existence of two or more long-term conditions in an individual

Non-communicable diseases (NCDs): also known as chronic diseases are not passed from person to person. They are of long duration and generally slow progression. The four main types of non-communicable diseases are cardiovascular diseases (like heart attacks and stroke), cancers, chronic respiratory diseases (such as chronic obstructed pulmonary disease and asthma) and diabetes.

Psychosis: A mental disorder (e.g., schizophrenia) characterized by delusional or disordered thinking detached from reality; symptoms often include hallucinations.

Schizophrenia: A psychotic disorder characterized by symptoms that fall into two categories: (1) positive symptoms, such as distortions in thoughts (delusions), perception (hallucinations), and language and thinking and (2) negative symptoms, such as flattened emotional responses and decreased goal-directed behavior.

In order to harness the power of emerging disease data, systems are needed to collect and transform this information in knowledge that could be analyzed, interpreted by doctors, and applied in order to improve patients life.

The ideal would be the integration of patient data at all levels, clinical, genomic and treatment data. Due to the limitations that we have to some data, this thesis will focus in developing a methology and tool that allows to examine the patient history data to obtain comorbidity patterns in the population. Due to the lack of patient genomic information in our databases, but taking into account the amount of knowledge driven every day in gene-disease association that is published every day, a gene-disease association database focus on psychiatric disorders has been also developed to understand the genetics underlying this disorders that affect a huge amount of population.

To efficiently conduct research to improve healthcare delivery and to improve the state of biomedicine by advancing its science. Informatics for integrating biology seeks to provide this instrumentation for using the informational by products of health care and the biological materials accumulated through the delivery of health care. (Isaac S Kohane, "A translational engine at the national scale: informatics for integrating biology and the bedside")

The main goal is to prove clinical investigators with the tools necessary to integrate medical record and clinical research data in the genomics age, a software suite to construct and integrate the modern clinical research chart.

#Deliverable Brunak

In the era of precision medicine, data-driven research is a natural supplement for hypothesis-driven research. Data-driven research will be a key player to discover evidence for further investigation, and provide knowledge on the whole range of diseases and conditions observed at clinics, as suggested by (Collins & Varmus 2015). The statement is especially true for disease occurring together less than expected, inverse comorbidities, which can help uncover protective mechanisms and thereby

enhance the understanding of diseases (Tabarés-Seisdedos & Rubenstein 2013; Ibáñez et al. 2014).

Previous cross-sectional data-driven studies have made use of the Medicare data, which consists of individuals 65 years or older in the period 1990-1993 (Hidalgo et al. 2009). The study was the first of its kind, but lacked rigorous statistics and only covered a very small period of time with a bias to geriatric disorders. To uncover novel disease relationships it is crucial to have a long term follow up across a full population. The Danish National Patient Registry has been collecting information on Danish hospital admissions using the ICD-10 coding system since 1994 across the whole population of 6.9 million Danes, and is considered one of the world's most complete registries ((Frank 2000; Lynge et al. 2011)). Novel disease relationships include both the co-occurrence of two diseases, comorbidity, and the occurrence of several morbidities in a single patient, multimorbidity. Separating well-known and novel relationships is demanding in itself. This is usually done by searching the literature, for example using PubMed. Other approaches have shown that an automated named-entity recognition system is also very efficient at picking up known associations between diseases and genes (Pletscher-Frankild et al. 2015). (Bravo et al. BeFree)

Diseases are driven by genetic and environmental factors. Some co-occurrences of diseases may be genetically founded, either due to a common risk loci (pleiotropic effects), or a disruption of the intra-cell protein-protein networks (Goh et al. 2007; Park et al. 2009). The studies have focused primarily on Mendelian disorders as defined in the Online Mendelian Inheritance in Man database (Hamosh et al. 2005). Complex disorders naturally do not have a single disruptive locus, but the genetic impact usually stem from a lot of smaller effects scattered in the genome. An in-depth understanding of the genes associated to disorders and the genetic interplay between them (and their interaction with the environment) will provide novel insights into both progression and manifestation.

##Deliverable Brunak

Biomedical research on the big data era can enable to predict the risk of suffering one disorder based in the other disorder you have, and prescribe the best treatment for the bunch of disorders suffered by the patient, walking through the personalized medicine.

Looking at the patients' medical history and ideally the genome information along with molecular data, doctors will be able to identify precisely what the patient is suffering from, and device a personalized treatment. The doctor will be able to predict the risk of suffering other diseases and if other family members will likely develop the same diseases.