Epidemiology and genetics of common mental disorders in the general population: the PEGASUS-Murcia project

Fernando Navarro-Mateu,1,2 MJ Tormo,2,3,4 G Vilagut,2,5 J Alonso,2,5,6 G Ruiz-Merino,7 T Escámez,7,8 D Salmerón,2,3,4 J Júdez,7,8 S Martínez,9 C Navarro2,3,4

ABSTRACT

Background: Multidisciplinary collaboration between clinicians, epidemiologists, neurogeneticists and statisticians on research projects has been encouraged to improve our knowledge of the complex mechanisms underlying the aetiology and burden of mental disorders. The PEGASUS-Murcia (Psychiatric Enquiry to General Population in Southeast Spain-Murcia) project was designed to assess the prevalence of common mental disorders and to identify the risk and protective factors, and it also included the collection of biological samples to study the gene–environmental interactions in the context of the World Mental Health Survey Initiative.

Methods and analysis: The PEGASUS-Murcia project is a new cross-sectional face-to-face interview survey based on a representative sample of non-institutionalised adults in the Region of Murcia (Mediterranean Southeast, Spain). Trained lay interviewers used the latest version of the computer-assisted personal interview of the Composite International Diagnostic Interview (CIDI 3.0) for use in Spain, specifically adapted for the project. Two biological samples of buccal mucosal epithelium will be collected from each interviewed participant, one for DNA extraction for genomic and epigenomic analyses and the other to obtain mRNA for gene expression quantification. Several quality control procedures will be implemented to assure the highest reliability and validity of the data. This article describes the rationale, sampling methods and questionnaire content as well as the laboratory methodology.

Ethics and dissemination: Informed consent will be obtained from all participants and a Regional Ethics Research Committee has approved the protocol. Results will be disseminated in peer-reviewed publications and presented at the national and international conferences.

Discussion: Cross-sectional studies, which combine detailed personal information with biological data, offer new and exciting opportunities to study the gene–environmental interactions in the aetiology of common mental disorders in representative samples of the general population. A collaborative multidisciplinary research approach offers the potential to advance our knowledge of the underlying complex interactions and this opens the field for further innovative study designs in psychiatric epidemiology.

Strengths and limitation of this study

- The assessment of environmental and genetic factors not only associated to mental disorder but also with positive mental health in a representative sample of the general population.
- A multidisciplinary research team better approaches the study of the complex interactions between environmental and genetic risk and protective factors involved in mental disorders.
- Its cross-sectional design which, while it allows association studies and the generation of new hypotheses, limits the possible causal interpretation of the findings.

BACKGROUND

The World Mental Health (WMH) Survey Initiative is a WHO initiative specifically designed to carry out the epidemiological surveys in a representative number of countries in all major regions of the world.1–3 All previous WMH surveys have used or are currently using the same diagnostic interview, the WHO Composite International Diagnostic Interview (WMH-CIDI, hereafter referred to as CIDI), a fully structured research diagnostic interview questionnaire designed to be used by trained lay interviewers without clinical experience. This initiative has generated an enormous body of comparative cross-national data on the epidemiology of mental disorders all over the world.5–7 As part of it, the European Study of the Epidemiology of Mental Disorders (ESEMeD) project was designed to collect data from representative samples of the adult population in six European countries: Belgium, France, Germany, Italy, the Netherlands and Spain.8 9 It has also generated a large number of scientific papers on the most prevalent mental health disorders (mood, anxiety and alcohol abuse) in Europe.10–17 There is a general consensus on the importance of the ESEMeD project in
Terms of improving scientific knowledge of the epidemiology of mental disorders in Europe.129

Genes and environment factors in the aetiology of mental disorders

Despite decades of intensive research, it remains difficult to identify specific genes and to characterise those environmental factors primarily responsible for mental disorders.18–22 The concept of genes and environmental factors as independent causes of mental disorders has been replaced by one of the complex interactions between them. These gene–environment (G×E) interactions imply a genetic predisposition of some subjects to be expressed differently depending on the environment to which they are exposed.23–24 For example, the important role of environmental factors, especially Stressful Life Events (SLEs), is now widely accepted. Exposure to various SLEs (work or physical problems, assault, natural disasters, etc), separately or cumulatively over the life of an individual, increases the risk of depression although in only a proportion of those exposed.25–26 These data suggest the existence of genetic differences which might explain individual variation in the sensitivity of people to the depressogenic effects of SLEs. On the other hand, the serotonin transporter (SERT or 5HTT) gene, a key regulator of serotonergic neurotransmission and one of the most studied genetic polymorphisms in relation to affective disorders,24 has been associated with depression,27–28 neuroticism,29 and post-traumatic stress disorder (PTSD).30 However, these findings have not always been replicated.31–33

These inconsistencies may be explained by, at least, three different factors. First, in adults, higher levels of neuroticism are associated with an increased risk of depression,34 anxiety30 and PTSD after exposure to a traumatic event,36 and are a powerful predictor of comorbidity between depression and anxiety.37 Neuroticism includes those personality traits that represent how some people perceive the world around them as threatening or stressful. In addition, some personality traits also influence the individual tendency to be potentially exposed to stressful environments. Predisposed individuals may tend to choose environments prone to having a high risk of exposure to stressful events. Specifically, this scenario, known as G×E correlation, may mediate the relationship between neuroticism and specific SLEs.38 Second, the genetic factors influencing the level of neuroticism, including the 5-HTTLPR polymorphism, are shared by persons having anxious-depressive spectrum disorders.37–38 Lastly, G×E interactions have been described involving 5-HTTLPR and depression,39 anxiety41 and PTSD.42 Despite all the above evidence, genetic association and G×E interaction studies do not usually analyse or control for the level of neuroticism in the relationship between 5-HTTLPR, SLEs and anxious-depressive spectrum disorders.

However, the question arising in this context is how environmental and genetic factors interact to produce a mental disorder.21–45 In recent years, increasing interest in the epigenetic factors described in other human diseases has focused on its role in mental disorders.46 The study of the epigenome, changes in gene expression by modulating the accessibility of information that occurs without modifying the DNA sequence, suggests that, although inheritable, these changes are not necessarily stable over the life span of individuals and can be modified under some environmental stimuli that modulate the activity of the enzymes involved, opening new prospects for developing therapeutic approaches based on epigenetic mechanisms.47 Epigenetic mechanisms have been associated with different mental disorders including depression,46 PTSD,47 schizophrenia,48–49 autism,48 bipolar disorder49 and alcohol dependence.50 In fact, epigenetic regulation of the glucocorticoid receptor signalling in neurons has been recently shown to be the mechanism underlying G×E interactions to explain the risk and resilience of PTSD after SLE in childhood.51

In order to integrate all these findings and create new opportunities and challenges offered by the G×E interaction scenarios in the field of mental disorders, a multidisciplinary collaboration between clinicians, epidemiologists, geneticists and statisticians offers greater opportunities.20–23,52 One of the proposed mechanisms for this collaboration includes carrying out community psychiatric surveys and this has been facilitated by the possibility of obtaining DNA and/or mRNA from peripheral tissues. Specifically, saliva or buccal cells offer an easy, saving, inexpensive and non-invasive method with accumulating scientific rationale to be added in general population surveys.53–56 The changes in gene expression can also be due to transcriptional alterations. In order to deepen the understanding of molecular mechanisms implicated in mental disorders, it is relevant to take into account transcriptional analyses with the RNA obtained at the same time as the DNA samples. The opportunity to get both biological samples at the same time from saliva offers the challenge of testing the suitability of this material for transcriptional analyses in general population surveys. Population-based surveys offer several advantages over other study designs to contribute to the clarification of the G×E interactions in mental disorders.43–59,60 First, the current knowledge of genes as risk factors is based almost exclusively on clinical and non-representative population samples. Second, the distribution of the gene polymorphisms of interest in the general population has not been well investigated. Third, this type of study can provide samples for future case–control studies and can be the bases for future longitudinal ones. Finally, hypotheses generated from epidemiological surveys may contribute to test new basic studies and can be considered as a complementary strategy to translational research.

Psychiatric Enquiry to General Population in Southeast Spain–Murcia project

Spain actively participated in the ESEMeD project with a representative sample of the adult general Spanish population (n=5473) and the results have been published in

Open Access
the national and the international journals. However, the sample size within most of the Autonomous Communities in Spain was too small to be able to achieve accurate and precise estimates at the regional level where healthcare policies are decided. Moreover, several differences between the Autonomous Communities in Spain in important aspects related to mental health such as socioeconomic and territorial inequalities in healthcare supply and in long-term care, access to and use of healthcare facilities, premature deaths due to alcohol consumption and the prevalence of psychological distress have recently been described.

Murcia is one of the 17 Autonomous Communities of Spain. It is a located in the southeast of the country on the Mediterranean coast, with a population of 1,424,063 inhabitants at the time of the survey (INE 2008, National Statistical Institute of Spain), almost one-third of them (30.7%) living in the capital.

The PEGASUS-Murcia (Psychiatric Enquiry to General Population in Southeast Spain-Murcia) project has been designed in order to obtain regional data of the prevalence, burden and care of a representative sample of the general adult population of Murcia in order to allow planning of new regional mental health policies and to compare the results with the national sample of Spain, Europe and all other countries participating in the WMH Survey Initiative. The project also constitutes a unique opportunity to initiate a biological bank of a well-studied representative sample of the general population.

**Objectives**

The PEGASUS-Murcia project is a multipurpose, observational, cross-sectional, comparative study of the non-institutionalised general population of Murcia region whose objective is to improve knowledge about common psychiatric disorders in two main areas. The first one is the epidemiology of mental disorders and protective and risk factors in the general population of Murcia. The specific objectives are: (1) to estimate the 1-month, 12-month and lifetime prevalence of the most common mental disorders, specifically mood and anxiety disorders, in the general population of Murcia; (2) to assess the independent association of mood and anxiety disorders with sociodemographic factors (gender, age, education and urban/rural location) and selected risk factors (family history, childhood experiences, religion, partnership status and sexual problems, among others); (3) to assess the quality of life of persons with the most common psychiatric disorders and to analyse how other variables (physical medical conditions and sociodemographic factors) may influence this outcome; (4) to assess the treatment for these disorders and to evaluate

---

**Figure 1** Flow chart of the PEGASUS-Murcia (Psychiatric Enquiry to General Population in Southeast Spain-Murcia) project. The response rate is defined as: (completed interviews)/(total released respondent sample cases–respondent non-sample cases). High-risk individuals: those who positively answer a number of specific questions related to mood and anxiety disorders in the screening section. Low-risk individuals: those without symptoms related to mood and anxiety disorders in the screening section. Long Path inclusion criteria: (a) all high-risk individuals and (b) a random subsample of 25% of the low-risk individuals. The remaining 75% of respondents without screening symptoms not randomly selected for the long path will follow the short path of the questionnaire.
the unmet need and the quality of care received and (5) to compare our results with those obtained from Spain, Europe and other non-European countries, including the USA. The second area is the genetic, epigenetic and transcriptomic influences associated with mental disorders. Its specific aims include (1) the estimation of the distribution of different candidate genes in the general population and their association with different psychiatric disorders; (2) the identification of sensitive alleles underlying potential G×E interactions and the study of epigenetic mechanisms involved, especially DNA methylation and (3) the analysis of gene expression alterations through transcriptomic assays.

METHODS AND ANALYSIS

Study design
The project is a cross-sectional face-to-face interview survey based on a representative sample of the adult and non-institutionalised general population of the Murcia region. Those who complete the interview will be invited to provide two biological samples from their oral mucous membranes. The target population is defined as persons aged 18 or older residing in Murcia, not living in institutions and with an active health card (defined as persons included in PERSAN, a regional registry that contains all residents with a health card which is periodically up-dated). The exclusion criteria are (1) Confirmed irretrievable contact errors (eg, telephone number and/or address); (2) institutionalised individuals (eg, in prison, in a hospital or in another institution) or those living outside the Autonomous Community during the survey field work and (3) individuals not able to understand the Spanish language or not able to conduct the questionnaire due to his/her physical or mental condition.

Sampling plan
The geographical area of the survey is the Murcia region, and a two-stage, stratified sampling design has been used. The primary sampling unit is the Primary Health Centre and the second is the individual. The sampling frame has been PERSAN, the regional healthcare population database in Murcia. Primary Health Centres have been grouped into nine strata, the current healthcare areas in Murcia region. The initial sample size was 4500 adult individuals divided into nine healthcare areas with proportionate allocation. A representative sample of two centres has been chosen in each health area, without individual participant replacement. Selection probability for each centre was known a priori and it was proportional to the size of the centre (% of adult individuals registered in the centre) and the proportion of adult individuals in the centre whose place of residence was rural, semiurban or urban. Within each of the two selected health centres, a stratified, random sample procedure, performed for each combination of gender, age group (18–24, 25–34, 35–49, 50–64 and 65+) and type of residence (rural, semiurban and urban), constitutes a stratum, and individuals have been selected using simple random sampling.

For each healthcare area, the sample size of each stratum has been selected such that the individuals with the same demographic characteristics had equal probability of being selected independently of the selected centre. If a high number of those fulfilling the exclusion criteria in one area is reached, a fixed number of additional individuals will be released (subsequent releases), according to the number of interviews completed in the area and following the same selection procedure within each of the centres as the ones used to select the initial release (no new centre will be selected for these releases). Any replacement of those persons who do not want to collaborate or who do not meet the non-eligibility criteria is not allowed.

Survey procedures and data control
Those selected will receive no financial incentive to participate and there will be no individual replacement procedure. Trained lay interviewers carry out the survey using the computer-assisted personal interview (CAPI) that was programmed centrally using the Blaise software system. This is an interviewing application developed by Statistics Netherlands (Herleen, the Netherlands) and designed to ease the handling of elaborate skip and complex randomisation patterns and to facilitate data entry, allow the elaboration of some questions and direct the interviewer through the questioning sequence.

Periodically, the completed interviews will be submitted to the Central Project Data Center (Regional Mental Health Service, Murcia, Spain) for checking and storage following a predetermined security procedure. All raw data will be transferred to the Hospital del Mar Medical Research Institute (IMIM) and the Department of Health Care Policy at Harvard University, coordinating centres of the ESEMeD and WMH Survey Initiative projects, respectively, via secure websites. The database has been declared to the Spanish Data Protection Agency.

A survey firm has been contracted to undertake the fieldwork and, in order to ensure the quality of the survey, several strategies are being implemented: (1) a 1-week training course for all interviewers by WHO-certified trainers on the original protocol and use of the CAPI version of the CIDI; (2) development of a written manual to standardise the interviewing procedure and all scientific and administrative elements that could affect comparability of data; (3) regular meetings with the survey firm to ensure adherence to the protocol and to deal with any difficulty that may have arisen and (4) data quality analysis to detect any inconsistencies and/or incomplete data.

The survey firm has been provided with sufficient data to allow contact with each of the individuals of the selected sample and only after 10 unsuccessful attempts the person will be considered as not-contactable or after confirmation that the selected person does not live at that address and new contact information is unavailable. Several methods will be used to improve the
participation of those selected: (1) an informative flyer providing general information related to the project and giving notice of future contact will be sent by conventional post together with an invitation letter signed by a person from the healthcare authority; (2) a phone call to invite them to participate in the interview process and to offer them the possibility to do the interview either at home or in their Primary Care Center; (3) several informative sessions for the healthcare personnel of the Primary Care Centers will be organised to facilitate their collaboration should the participants ask them about the project; (4) during the period when the interviews will take place, some official posters will be put in public centres to inform the people about the project; (5) all interviewers will be provided with an official identification and have been trained on how to explain the institutional nature of the research project.

Survey questionnaire
The questionnaire used in the PEGASUS-Murcia project is a revised version of the CIDI which, together with diagnostic information on the most common mental disorders, also includes specific information on the severity of the disorders, symptoms, disability, quality of life, use of services and medication and several risk factors.

Composite International Diagnostic Interview
The Composite International Diagnostic Interview (CIDI) is a comprehensive, highly structured interview specifically designed by the WHO for the purpose of ascertaining diagnoses of mental illnesses based on the WHO International Classification of Disease (ICD-10) and not exclusively on DSM definitions and criteria. This objective is particularly important for cross-national comparative research of the epidemiology of mental illnesses throughout the entire world. It comprises nearly 5000 questions divided into 42 sections (table 1) and these, in turn, are grouped into two main parts: diagnostic and other. The first includes the clinical part of the interview with an introductory screening section and 22 diagnostic sections that assess different psychiatric conditions. The second includes various non-clinical sections that assess utilisation of services, use of psychotropic drugs, degree of functioning in several aspects, chronic physical conditions, risk factors, social networks, caregiver burden and sociodemographic variables.

The most recent version of the CIDI (V.3.0) is the end result of a number of international studies and adaptations made since 2000 when it was first used in WMH surveys. It was first created in English and has been translated into more than 30 different languages using the standard WHO protocol with a rigorous process of adaptation. Several clinical reappraisal studies have been carried out and the concordance of the CIDI V.3.0 has been evaluated in different subgroups of WMH surveys using the Structured Clinical Interview for DSM-IV (SCID) as the clinical gold standard and a moderate-to-excellent concordance has been found for most mental disorders. CIDI is available in two formats: the paper form or PAPI (Paper and Pencil Interviewing) and the computerised form or CAPI, designed to ease the handling of elaborate skip and complex randomisation patterns and to facilitate data entry with a resulting reduction in interview time and errors in data collection and recording. The original Spanish CAPI version used in Spain had not been updated since it was used in the context of the ESEMeD project almost 10 years ago. Since then, all improvements in the questionnaire have only been added to the CIDI Latin American V.20.0. However, due to the linguistic and cultural differences in Spanish-speaking populations, this CAPI version had to be culturally adapted for use in Spain by our research team and this process is fully described elsewhere.

To further shorten the length of the questionnaire, some sections were not selected for the purposes of this project. These include Intermittent Explosive Disorder, Personality I and II, Neuropsychiatric and Pre-Menstrual and Gambling sections. Some others were substituted by other questions or questionnaires, for example, the Tobacco Use section was simplified using some questions obtained from the Spanish National Health Survey and the Psychosis section with the Community Assessment of Psychic Experiences instrument (CAPE), both described below.

Other study instruments
Several other instruments were added to the original CIDI for the specific purposes of the PEGASUS-Murcia project. These include the Spanish version of different questionnaires: (1) Mini-Mental State Examination for interviewees older than 60 years, (2) the Cognitive Failure Questionnaire, (3) the Neuroticism, Extroversion and Lie subscales of the abbreviated version of the Eysenck Personality Questionnaire (EPQR-A), (4) the Resilience Scale, (5) the CAPE to measure attenuated psychotic symptoms in the general population instead of the Psychosis section of the CIDI, as the latest is only used as a screening instrument in the detection of psychosis. Those who positively answer two items of the positive dimension with a score equal or superior to 3, have been hospitalised for psychiatric reasons and/or have received psychotropic medication during the last year will be evaluated by a clinic psychiatrist with the module C (psychotic disorders) of the SCID; (6) a brief list of 12 SLEs in the last 12 months was included by the combination of a list of threatening experiences and the emotional and life-changing impact of each event; (7) the European Quality of Life Scale (EuroQol 5D) and the Short Form 12 Health Questionnaire (SF-12 v2); (8) an ad hoc questionnaire of partner violence obtained from the Spanish National Health Survey and from the regional mental health clinical guidelines and (9) finally, some questions related to tobacco use and physical exercises from the Spanish National Health Survey.
**Questionnaire pathways**

In order to optimise the duration of the interview, the WMH questionnaire was divided into two parts with questions in part 1 administered to all respondents and those in part 2 only to a subsample of individuals who followed the long path of the interview. Part 2 of the interview includes detailed information about a wide range of aspects related to the primary disorders and also to mental disorders of secondary interest (table 1). The inclusion criteria for the long path are (1) all individuals that could be considered as ‘high-risk individuals’ because they positively answer a number of specific questions related to mood and anxiety disorders and (2) a random subsample (25%) of the respondent without

### Table 1 Description of the adapted version of the WHO-Composite International Diagnostic Interview (WHO-CIDI) used in the PEGASUS-Murcia project

<table>
<thead>
<tr>
<th>Sections</th>
<th>Module</th>
<th>Number of items</th>
<th>Rules for administration*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Household listing</td>
<td>Methodological</td>
<td>5</td>
<td>All respondents</td>
</tr>
<tr>
<td>screening (SCR)</td>
<td>Screening</td>
<td>51</td>
<td>All respondents</td>
</tr>
<tr>
<td>Minimental state examination</td>
<td>Risk factors</td>
<td>106</td>
<td>All respondents</td>
</tr>
<tr>
<td>Quality/lie subscale‡</td>
<td>Functioning and physical disorder</td>
<td>106</td>
<td>All respondents</td>
</tr>
<tr>
<td>Depression</td>
<td>Mood disorder</td>
<td>143</td>
<td>All respondents</td>
</tr>
<tr>
<td>Mania</td>
<td>Mood disorder</td>
<td>16</td>
<td>All respondents</td>
</tr>
<tr>
<td>Panic disorder</td>
<td>Anxiety</td>
<td>85</td>
<td>All respondents</td>
</tr>
<tr>
<td>Specific phobia</td>
<td>Anxiety</td>
<td>51</td>
<td>All respondents</td>
</tr>
<tr>
<td>Social phobia</td>
<td>Anxiety</td>
<td>51</td>
<td>All respondents</td>
</tr>
<tr>
<td>Agoraphobia</td>
<td>Anxiety</td>
<td>51</td>
<td>All respondents</td>
</tr>
<tr>
<td>General anxiety disorder</td>
<td>Anxiety</td>
<td>51</td>
<td>All respondents</td>
</tr>
<tr>
<td>Suicidality</td>
<td>Other diagnostic</td>
<td>51</td>
<td>All respondents</td>
</tr>
<tr>
<td>Use of services</td>
<td>Treatment</td>
<td>51</td>
<td>All respondents</td>
</tr>
<tr>
<td>Group of questions (tobacco and physical exercise)</td>
<td>Risk/protective factors</td>
<td>22–32</td>
<td>All respondents</td>
</tr>
<tr>
<td>Pharmacoepidemiology</td>
<td>Treatment</td>
<td>241</td>
<td>All respondents</td>
</tr>
<tr>
<td>Substances</td>
<td>Substance abuse</td>
<td>46</td>
<td>All respondents</td>
</tr>
<tr>
<td>Post-traumatic stress disorder</td>
<td>Anxiety</td>
<td>46</td>
<td>All respondents</td>
</tr>
<tr>
<td>Chronic conditions</td>
<td>Functioning and physical disorder</td>
<td>46</td>
<td>All respondents</td>
</tr>
<tr>
<td>30 days functions</td>
<td>Functioning and physical disorder</td>
<td>46</td>
<td>All respondents</td>
</tr>
<tr>
<td>30 days symptoms</td>
<td>Functioning and physical disorder</td>
<td>46</td>
<td>All respondents</td>
</tr>
<tr>
<td>Eating disorders</td>
<td>Other diagnostic</td>
<td>46</td>
<td>All respondents</td>
</tr>
<tr>
<td>Obsessive-compulsive disorder</td>
<td>Anxiety</td>
<td>46</td>
<td>All respondents</td>
</tr>
<tr>
<td>CAPE</td>
<td>Psychosis</td>
<td>46</td>
<td>All respondents</td>
</tr>
<tr>
<td>CFQ</td>
<td>Risk factors</td>
<td>46</td>
<td>All respondents</td>
</tr>
<tr>
<td>SLE</td>
<td>Risk factors</td>
<td>46</td>
<td>All respondents</td>
</tr>
<tr>
<td>Neuroticism and extroversion subscales‡</td>
<td>Risk/protective factors</td>
<td>46</td>
<td>All respondents</td>
</tr>
<tr>
<td>Resilience scale</td>
<td>Protective factors</td>
<td>46</td>
<td>All respondents</td>
</tr>
<tr>
<td>Employment</td>
<td>Sociodemographics</td>
<td>46</td>
<td>All respondents</td>
</tr>
<tr>
<td>finances</td>
<td>Sociodemographics</td>
<td>46</td>
<td>All respondents</td>
</tr>
<tr>
<td>Marriage</td>
<td>Sociodemographics</td>
<td>46</td>
<td>All respondents</td>
</tr>
<tr>
<td>Partner violence</td>
<td>Risk factors</td>
<td>46</td>
<td>All respondents</td>
</tr>
<tr>
<td>Children</td>
<td>Sociodemographics</td>
<td>46</td>
<td>All respondents</td>
</tr>
<tr>
<td>Social networks</td>
<td>Risk/protective factors</td>
<td>46</td>
<td>All respondents</td>
</tr>
<tr>
<td>Adult demographics</td>
<td>Sociodemographics</td>
<td>46</td>
<td>All respondents</td>
</tr>
<tr>
<td>Child demographics</td>
<td>Sociodemographics</td>
<td>46</td>
<td>All respondents</td>
</tr>
<tr>
<td>Demographic short</td>
<td>Sociodemographics</td>
<td>46</td>
<td>All respondents</td>
</tr>
<tr>
<td>Childhood</td>
<td>Sociodemographics</td>
<td>46</td>
<td>All respondents</td>
</tr>
<tr>
<td>Attention hyperactivity</td>
<td>Childhood</td>
<td>46</td>
<td>All respondents</td>
</tr>
<tr>
<td>Oppositional defiant</td>
<td>Childhood</td>
<td>46</td>
<td>All respondents</td>
</tr>
<tr>
<td>Conduct disorder</td>
<td>Childhood</td>
<td>46</td>
<td>All respondents</td>
</tr>
</tbody>
</table>

Continued
symptoms (‘low-risk individuals’). The remaining 75% of respondents without screening symptoms not randomly selected for the long path followed the short path. The computer, without any intervention of the interviewer, automatically makes all these pathways. In this shorter itinerary, a specific section that included those questions needed to calculate some demographic indicators substituted the sections omitted. Moreover, two sections were only used in a percentage of the long path itinerary, eating disorders (50%) and obsessive-compulsive disorder (33%).

Quality control procedures

Data quality will be controlled in a number of ways to ensure that the predetermined protocol has been followed achieving the greatest reliability and validity and these quality control procedures will be organised and supervised by the members of the coordinating centres. The principal investigator will review all responses to open-ended questions to check whether the narratives exclude a clinical diagnosis of mental disorders, that is, whether the symptoms were due to a physical illness. All these procedures will be verified by the coordinating centres and the final document included several aspects, for example, sample releases, the duration of the interview and the proportion of positive responses to selected screening questions. Local members of the research team will be responsible for verifying the informed consent forms and the quality checking following computerised protocols. These procedures are similar to those implemented in the ESEMeD project and are fully described elsewhere. Briefly, they consist of checks of individual pieces of information from the interviewees, for example, completion status, consistency across the questionnaire, questionnaire itinerary and length of the interview, and from the interviewers, number of disorders screened positively, verification of a random selection of almost 1% of interviews completed by a telephone contact to confirm the interview and some aspects related to it such as place, approximate duration and identification of the interviewer.

Laboratory methods

On completion of the interview, interviewees will be asked to provide two biological samples of buccal mucosal epithelium, one for DNA extraction for genomic and epigenomic analysis and the other to obtain mRNA for gene expression quantification (transcriptomic assays). These samples will be obtained only if the interviewee signs informed consents specifically designed for this project based on international recommendations for population-based research involving genetics and previously approved by the Regional Ethics Research Committee. Interviewers have been trained by one of the authors (TE) to adequately obtain the biological sample by scraping the oral mucosa using swabs compatible with molecular amplification techniques, as they do not interfere with the amplification process (FLOQSwabs Flocked Swabs, Copan Flock Technologies srl).

Samples for DNA extraction will be collected in sterile 1.5 mL tubes. Those to be used for RNA extraction will be harvested in dark sterile tubes containing RNA protect cell reagent (QIAGEN, Hilden, Germany), which provides immediate stabilisation of RNA. Cells will be thus stabilised at room temperature and can then be stored or transported at ambient temperature prior to RNA purification. Tubes will be labelled with tags (14C. B. 40×40 type) with a specific code for each sample and will be packaged and sent to BIOBANC-MUR (the biobank for biomedical research network of the Region of Murcia, RD09/0076/00065, as a partner of the Spanish National Biobanks Network; IMIB: Instituto Murciano de Investigación Biosanitaria) according to the current Spanish legislation and following the regulations of the International Air Transport Association (IATA) on biological sample shipping.

Those sample accepted by BIOBANC-MUR will be registered using a specific biobanking software (bio-e-bank, VITROSOFT, SL), as part of a Laboratory Integrated

<table>
<thead>
<tr>
<th>Table 1 Continued</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sections</strong></td>
</tr>
<tr>
<td>Separation anxiety disorder</td>
</tr>
<tr>
<td>Family burden</td>
</tr>
<tr>
<td>Quality/lie subscale†</td>
</tr>
<tr>
<td>Respondent contacts</td>
</tr>
<tr>
<td>Interviewer observation</td>
</tr>
</tbody>
</table>

*Long path inclusion criteria: (1) all individuals who could be considered as ‘high-risk individuals’, because they had positively answered a number of specific questions related to mood and anxiety disorders, and (b) a random subsample (25%) of the respondents without symptoms (‘low-risk individuals’). The remaining 75% of respondents without screening symptoms not randomly selected for the long path followed the short path of the questionnaire.

†Neuroticism and extraversion subscales of the abbreviated version of the Eysenck Personality Questionnaire (EPQR-A).

‡Lie subscale of the abbreviated version of the EPQR-A; PEGASUS-Murcia, Psychiatric Enquiry to General Population in Southeast Spain-Murcia.

CAPE, Community Assessment of Psychiatric Experiences; CFQ, Cognitive Failure Questionnaire; EQ-5D, European Quality of Life Scale; SF-12 v2, Short Form 12 Health Questionnaire; SLE, Stressful Life Events.
Management System (LIMS). The nucleic acid extraction will be performed automatically (QIAcube system; Qiagen, Hilden, Germany) to minimise the variability due to manual handling using QIAamp DNA Blood Mini Kit and RNeasyPlus Mini Kit (Qiagen, Hilden, Germany) for DNA and RNA extraction, respectively.

QIAamp DNA Blood Mini Kit provides fast and easy method for purification of total DNA for reliable PCR and Southern blotting from whole human blood,uffy coat, cultured cells, lymphocytes, plasma, serum, body fluids and buccal swabs. The synthesis of complementary DNA (cDNA) from mRNA for expression studies will be developed for all samples by reverse transcription using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems). All processes will be performed according to the manufacturer’s instructions.

Nucleic acids quantity and quality will be determined by the ratio A260/280 calculated based on 260 and 280 nm absorbance measured using a spectrophotometer.94–96 The ratio A260/230 is commonly used as a secondary indicator of nucleic acid purity.97–99 The integrity of DNA will be visualised by electrophoresis on 1% agarose gel (migration for 1 h at 100 V) using 100 ng of total DNA and a 23 kb DNA ladder (Lambda DNA/HindIII Marker (Thermo Fisher Scientific) as DNA marker. All mRNA samples will be transformed into cDNA.

Specially trained technicians from the BIOBANC-MUR will be used to monitor the specimen collection by donors and to perform sample manipulations in order to minimise the variability of results and to obtain the optimal quality of nucleic acids for this and future studies. The processed biospecimens (150 μL of DNA and 80 μL of cDNA) will be stored in 750 μL microtubes in an ultra-freezer at −80°C located in BIOBANC-MUR.

Statistical methods

The expected response rate (RR) has been set to a minimum of 65%, based on a previous regional community survey which included the donation of blood samples.100 101 The RR will be calculated based on the proportion of people interviewed and was defined as the number of completed interviews divided by the total number of cases minus the number of non-eligible cases.

Weighting procedures

Given that the interview is divided into two parts and only a portion of the sample will be selected for the second part, two types of weightings are considered to estimate population parameters. The first is to weight for the probability of selection for each healthcare area, health centre and demographic stratum and the second is for the random skips included in the questionnaire. The method designed is described in box 1.

Analysis of the data and forthcoming research projects

There are three data analysis centres in the project: Harvard University (Boston, USA), IMIM (Barcelona, Spain) and the Regional Centers of Epidemiology and Mental Health (Murcia, Spain). Harvard will supervise all quality procedures and provides consultancy in many aspects of the analysis, including the sampling design, the weighting procedures and the verification of the CIDI diagnostic algorithms. All the analyses will be performed using SAS and SPSS programs.

Related to this research project, several other lines of research with different designs are being developed, for example, case-control studies and meta-analyses. An example of the former is a case-control study of the G×E interactions, involving 5-HTTLPR polymorphisms, located in an area where a recent earthquake took place in Lorca (Murcia). It has been specifically designed to analyse its impact in the mental health of the general population exposed. Cases will be those people with a diagnostic of affective and/or anxiety disorder exposed to the earthquake attended in the Mental Health Care Centre and controls will be obtained from those exposed to the earthquake that are going to be interviewed in the PEGASUS-Murcia project and without a diagnosis of any affective and/or anxiety disorder. Recently, our research team has published a meta-analysis of the relationship between 5-HTTLPR polymorphism and PTSD.33

ETHICS AND DISSEMINATION

Eligible individuals will be asked to sign two independent informed consents to participate, the first one to be interviewed, including the possibility of future new contacts and the second to provide the biological samples but only those who had already completed the questionnaire. Name and contact information will be stored separately from any information provided as part of the study questionnaire. The Clinical Research Ethics Committee of the University Hospital Virgen de la Arrixaca of Murcia approved the protocol and the database of personal information has been registered with the National Data Protection Agency. Data from PEGASUS-Murcia project will be included in the WMH Cross National Sample for international comparisons. The study findings will be submitted to peer-reviewed journals for publication, and presented at the national and the international scientific meetings.

DISCUSSION

The epidemiology of mental illnesses is a fascinating but highly complex area of research. This complexity is primarily due to a wide range of factors, environmental and genetic, which combine to produce a recognised psychiatric disorder. Previous epidemiological research has resulted in the production of a great amount of data but it has been difficult to make cross-national comparisons due to methodological variability. The WMH

8

### Box 1 Weighting procedures

#### First weighting procedure:

Step 1: For each healthcare area \( h \), health centre \( c \) and demographic stratum (sex, age group and type of residence), all individuals have sampling weight \( w_i = 1/p_{hcgr} \), where \( p_{hcgr} \) is the probability that the centre \( c \) was selected, \( p_{hcgr} = n_{hcgr}/N_{hcgr} \) and \( n_{hcgr} \) is the sample size for the demographic stratum with \( N_{hcgr} \) individuals registered in the sampling frame.

Step 2: Non-response weight \( (w_{nr}) \) if \( p^*_{hcgr} \) is the proportion of eligible persons that is actually interviewed in the healthcare area \( h \), centre \( c \), sex \( s \), age group \( g \) and type of residence \( r \), the non-response weight of the persons in the healthcare area \( h \), centre \( c \), sex \( s \), age group \( g \) and type of residence \( r \) is

\[
w_{nr} = 1/p_{hcgr}.
\]

Step 3: Unadjusted weight \( (w_{unadj}) \) — it was calculated as the product of sampling weight by non-response weight:

\[
w_{unadj} = w_iw_{nr}.
\]

Step 4: Poststratification weight \( (w_{psk}) \) — data on population of the region of Murcia by sex, age and healthcare area were provided by the CREM (Centro Regional de Estadística de Murcia; Padrón 2010; http://www.carm.es/ecomun/sicrem/PU_padr0n/).

The population for the age group 18–24 has been estimated as the population for the age group 18–19 plus the population for the age group 20–24. The population for the age group 18–19 has been estimated as the population for the age group 15–19 times the proportion of population aged 18–19 in the age group 15–19 in Murcia: 0.4116 for males and 0.4165 for females. A poststratification weight was created to ensure that the joint distribution of the poststratifying variables healthcare area, sex and age group matches the known population joint distribution of Murcia.

Step 5: Adjusted weight \( (w_{adj}) \) — the adjusted weight of an individual in the healthcare area \( h \), centre \( c \), sex \( s \), age group \( g \) and type of residence \( r \) is

\[
w_{adj} = w_{unadj} \times w_{psk}.
\]

Step 6: Normalised weight:

\[
w_{norm} = w_{adj} n_{i=1} w_{adj}.
\]

Step 7: Trimmed weight \( (w_{trim}) \) — trim the normalised weight obtained from step 6. The upper and lower 5% were trimmed to the mean of each tail.

Step 8: Normalised trimmed weight:

\[
w = w_{trim} n_{i=1} w_{trim}.
\]

#### Second weighting procedure:

To take into account the random skips in the Composite International Diagnostic Interview questionnaire applied to define the long path we calculated the skip pattern weights. Only a portion of the sample completed the second part (part 2) of the survey. The probability of inclusion into part 2 is based on the presence or absence of disorder symptoms as defined in the interview schedule. Again, different steps will be followed:

Step 1: Part 2 selection weight \( (w_{psk}) \) — each individual \( i \) in the sample that accepted to respond the first part of the survey was selected into part 2 with probability \( \pi_i \) where \( \pi_i=1 \) for high-risk individuals of having mental disorders and \( \pi_i=0.25 \) for the rest. Then the part 2 selection weight of individual \( i \) is \( w_{psk}=1/\pi_i \).

### Box 1 Continued

Step 2: Unadjusted part 2 weight \( (w_{psk}) \) — the product of \( w_{psk} \) (part 1) and the part 2 selection weights.

Step 3: Part 2 poststratification weight \( (w_{pspk}) \) — similar to the previous poststratification procedure, a poststratification weight was created to ensure that the joint distribution of the variables healthcare area, sex and age group in part 2 match the known population distribution of Murcia.

Step 4: Part 2 adjusted weight \( (w_{psadj}) \) — the adjusted weight of an individual \( i \) in the healthcare area \( h \), centre \( c \), sex \( s \), age group \( g \) and type of residence \( r \) is

\[
w_{psadj} = w_{pspk} \times w_{psadj}.
\]

Step 5: Part 2 normalised weight:

\[
w_{psnorm} = w_{psadj} n_{i=1} w_{psadj}.
\]

Survey Initiative aimed to address this issue by using an international standardised protocol, allowing comparisons of the most common mental disorders and their associated factors throughout the world. Using this study design, it therefore offers the opportunity for new surveys to be performed in the context of an international collaborative initiative and the possibility to adapt the questionnaire according to the specific aims of the research being undertaken. The PEGASUS-Murcia project can be considered as an example of how the latter has been successfully achieved. It is a cross-sectional study designed to assess the prevalence of the most frequent mental disorders and their correlates in a representative sample of the general population of Murcia. Its primary strengths are: (1) the fact that it was specifically adapted to assess the factors not only associated with mental disorders but also with positive mental health in a representative sample of the general population; (2) its context focused on regional needs where healthcare decisions are taken regarding resource allocation and mental health planning; (3) the collection of biological samples not only for DNA analysis but also for mRNA; (4) all the information collected in our study, including biological samples, can be correlated with past and future health events because all Spanish population had free access to the healthcare system at the time of its inception and were thus registered and provided with a unique identification number, and therefore (5) finally, the inclusion of a multidisciplinary research team is in accordance with the international consensus regarding the need for interdisciplinary collaboration between clinicians, epidemiologists and neuroscience researchers to increase their combined efforts to study the complex gene–gene and G×E interactions underlying mental health disorders.23 60 102 103 Concerns have been expressed about the cost-effectiveness of psychiatric epidemiological surveys, such as WMH-2000 projects,104 an example being the rationale for starting a new psychiatric epidemiological survey in the Autonomous Community of Murcia if Spain had already participated in the ESEMeD project. However,
there are several reasons to justify this regional initiative. First, public health and healthcare agencies usually allocate mental health resources, including human, based on data from the national epidemiological surveys, such as that provided by the Spanish participation in the ESEMeD project. As previously mentioned, the involvement of the region of Murcia in the Spanish ESEMeD survey did not allow the evaluation of specific regional data. Nowadays, the main responsibility for planning and management of healthcare resources in Spain lies with the Autonomous Communities and differences exist between them in terms of accessibility, amount of healthcare resources and political decision-making. Devolution of this responsibility to Murcia occurred in December 2001.

Second, the inclusion of biological data in a well-designed multidisciplinary epidemiological study offers great advantages in terms of a more global understanding of mental disorders. These are complex illnesses of the brain where social, familial, psychological and biological elements interact throughout the entire life of a person to influence his/her risk of developing a mental health disorder. To extend our understanding of the physiopathology and epidemiology of the more common ones (mood and anxiety), it is necessary to identify the genetic loci and polymorphic alleles and their distribution in the healthy and affected population whose function in determining risk for, and protection against, these conditions probably depends on gene–gene and GxE interactions. The collection of genetic material from representative samples from the general population, well described using international diagnostic instruments such as CIDI, offers new and different possibilities to evaluate candidate genes in non-biased samples and to describe their distribution in the general population that may contribute to the clarification of the complexity of mental disorders.

Third, our project involving a multidisciplinary research team gives new opportunities to develop different study designs that can move from descriptive to analytical epidemiology. For example, this representative sample constitutes a good source of controls for future case-control studies, where cases will be provided from the public healthcare clinics, and can be the starting point for future cohort studies. Our project was designed to allow for all these possibilities.

LIMITATIONS OF THE STUDY
Currently, the main limitations of the PEGASUS-Murcia project are related to: (1) the cross-sectional design which, while it allows association studies, limits the possible causal interpretation of the findings. However, these findings may provide new hypotheses and enable the design of new studies; (2) not all interviewees will provide biological samples and this may affect the representativeness of some mental disorders in future analyses. To determine whether this will result in selection bias, we will analyse whether there are distinguishing characteristics between donors and non-donors in the distribution of mental disorders and other characteristics of the participants; (3) the population stratification in our study which will be used for future genetic association analyses is performed by using the stated ancestral origin by participants instead of using genetic markers and (4) biological samples will be obtained from oral mucosal scrapings and not from the brain neurons. However, this is a general situation given the ethical issues and difficulties in obtaining neural tissues and, in any case, gene expression does not appear to be specific to neural tissue, at least in some genes that have ubiquitous expression, for example, 5-HTTLPR.

CONCLUSIONS AND FUTURE DIRECTIONS
The PEGASUS-Murcia project is a sound base for multidisciplinary collaborative mental health research studies which will provide not only a huge amount of epidemiological information but will also offer exciting opportunities to clarify the complex interactions between genetic and environmental factors which result in a range of mental health disorders.

Author affiliations
1Unidad de Docencia, Investigación y Formación en Salud Mental (UDIF-SM), Subdirección General de Salud Mental y Asistencia Psiquiátrica, Servicio Murciano de Salud, Murcia, Spain
2CIBER de Epidemiología y Salud Pública (CIBERESP), Murcia, Spain
3Servicio de Epidemiología, Consejería de Sanidad y Política Social, Murcia, Spain
4Departamento de Ciencias Sociosanitarias, Universidad de Murcia, Murcia, Spain
5IMIM-Institut Hospital del Mar d’Investigacions Mèdiques, Barcelona, Spain
6Departamento de Salud y Ciencias Experimentales, Universidad Pompeu Fabra, Barcelona, Spain
7Fundación para la Formación e Investigación Sanitarias (FFIS) de la Región de Murcia, Murcia, Spain
8IMIB BIOBANC-MUR, Biobanco-HUVA-AECC-FFIS, Murcia, Spain
9Instituto de Neurociencias, UMH-CSIC, Alicante, Spain

Acknowledgements The authors would like to thank Carlos Giribert Muñoz, Deputy Director of Mental Health and Psychiatric Services of Murcia for his support in developing the PEGASUS-Murcia project; Inés Morán-Sánchez, Mª Luisa Pujalte and Ascensión Garriga for their contribution to the initial phases of this project; Monica Ballesta Ruíz for her contribution to the sample selection procedure; David Martínez Martínez for his contribution to the management of the software; all the collaborators from the BIOBANC-MUR (R Martín Marin, B Veas-Pérez de Tudela López, A Parra Montoya, E Sánchez Baeza); Pedro J Bernal for his collaboration with the PERSAN database; and, finally, to Mike Tobin for his helpful discussions and contribution during the English translation of the document. They also thank the WMH Coordinating Center staff at Harvard and Michigan Universities and, especially Professor Ron Kessler, for their assistance with the instrumentation, fieldwork and data analysis.

Contributors FN-M, MJT, GV, JA, TE, SM and CN conceived the design and supervised the whole process of the study. GV, JA and FM-M have coordinated the project with the WMH Survey Initiative. MJT, JA and CN are coordinating the epidemiological aspects. TE, JJ and SM are responsible for the genetic aspects. MJT, DS and GV were responsible for the sampling methods. GV, GR-M and DS are responsible of the implementation of the qualitative procedures and the statistical analyses. All authors read and approved the final manuscript.
Funding The PEGASUS-Murcia project is supported by the Regional Health Authorities of Murcia (Servicio Murciano de Salud and Consejería de Sanidad y Política Social) (Decreto no 455/2009), the Fundación para la Formación e Investigación Sanitarias (FFIS) de la Región de Murcia (No Expedientes: CM0829 I and FFIS/EMER09/14) and the Ayudas para proyectos de Investigación en Salud—ISCIII—del Plan Nacional de Investigación Científica, Desarrollo e Innovación Tecnológica (PI12/00809). The PEGASUS-Murcia project is carried out in conjunction with the WHO/MHM Survey Initiative. These activities were supported by the US National Institute of Mental Health (R01MH070884), the John D and Catherine T MacArthur Foundation, the Pfizer Foundation, the U.S. Public Health Service (R13-MH66849, R01-MH068664 and R01 DA016558), the Fogarty International Center (FIRCA R03—TW006481), the Pan American Health Organization, the Eli Lilly & Company Foundation, Ortho-McNeil Pharmaceutical, Inc, GlaxoSmithKline, Bristol-Myers Squibb and Shire. A complete list of WMH publications can be found at http://www.hcp.med.harvard.edu/wmh/.

Competing interests None.

Ethics approval Clinical Research Ethics Committee of the University Hospital Virgen de la Arrixaca of Murcia.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement Further details of the study protocol can be requested from the corresponding author.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/3.0/

REFERENCES


