Synopsis no.: S2.9

Preliminary title:
Incidence of psychotic disorders across Europe: analyses of determinants at small-area level in the EU-GEI study

Contact info for the person(s) proposing the synopsis

Name: **Andrei Szoke**

Partner no: **9/INSERM**

e-mail address: **andrei.szoke@inserm.fr**

Publication category: 2

Working and writing group:

A. Szoke, J. Kirkbride, I. Tarricone, D. Berardi, A. Lasalvia, C. Morgan, R. Murray, B. Rutten, J. van Os

Work Packages involved: WP2

Partners involved from whom candidate co-authors (*additional to working and writing group*) should be nominated:
All partners/sites involved in WP2 provided they have data with sufficient detail to be included in the analysis (see below)

Objectives (scientific background, hypothesis, methods, and expected results):

1. Scientific Background/ objectives

Epidemiological studies linking the spatial distribution of cases with area level characteristics are important for theoretical (i.e. identification of risk factors) and pragmatic reasons (i.e. mental health/ prevention policies, distribution of resources according to expected number of cases etc.)
Studies conducted to date have shown repeatedly that at small-area level (neighbourhood, electoral wards, postcodes etc.) significant differences exist in the incidence of psychotic disorders, especially schizophrenia.
Some of these studies explored potential risk factors at population level and found that several such variables (e.g. social fragmentation, deprivation, ethnic density, social capital) were correlated with risk of psychosis.
However, the number of such studies is small, data available originate from a very small number of countries (essentially the UK and Northern European countries) and the methodologies used were not standardized thus making comparisons and generalization of findings difficult.
To answer the limitations of our knowledge in this area our objectives are:
1. to investigate, at small-area level, variation in the incidence of affective, non-affective and total psychoses;
2. to study the factors that influence this variation
3. to compare the role/ weight of the different (potential) risk factors across different national contexts.

2. Hypotheses

1. Significant variation in standardized incidence is expected for non-affective psychoses but not for affective

psychoses;
2. Several risk factors will significantly influence distribution of non-affective disorders
3. At least for some of the risk factors at area level (e.g. ethnic density) the weight will be modified by the national context (significant interaction)

3. Methods

1. Centres included:
All centres will be included provided that:

a. the method used to identify cases ensures that data are not biased at (small area) geographical level;
b. each case can be assigned to a (small) geographic area according to the place of residence;
c. denominator and population risk factor data are available for the same area level

2. Area definition:
The period of time to collect cases is different between centres. Also, the degree of geographical precision (i.e. the size of the populations/areas) is expected to show variation between centres. We will aim to define area sizes across centres (for example by grouping areas in the centres where smaller area are available) in order to have similar persons*year samples across different centres. Also, for statistical reasons, we will choose the size of samples from each area in order to minimize the number of samples with 0 cases and, at the same time, to have enough geographical units (areas) in each centre.

3. Dependent variables
Incidence rates (indirect standardized for age and gender) for all included centres (separately for non-affective, affective and global psychoses)

4. Explanatory variables
4.1. Measures for economic deprivation, ethnic density and social disorganization. The choice of the exact measures (and between a single measure or an aggregate one) will depend on face validity and, more important, on availability of identical/similar data across the different national censuses.

4.2. Centres' characteristics (i.e. rural vs urban and country)

5. Statistical analyses
see below

Data needed for the study:

For each case (from partners): basic demographics (age/ gender), diagnosis (affective/ non-affective psychosis), geo-localisation (area)

For each area (from national censuses):
- denominator data (number of at-risk persons, age and gender structure);
- data on independent variables (putative risk factors): to be decided (see above)

For Bayesian/ spatial modelling a contiguity matrix (based on the spatial relationship between the different areas)

Plan for statistical analysis (overall strategy):

Based on available data from participating centres we will employ multilevel modelling, including Bayesian hierarchical models of spatial dependency, to investigate whether the incidence of psychotic disorders varies in each centre between small areas. Our modelling strategy will first investigate the extent of variation in incidence rates at the neighbourhood level within each centre, in order to compare this between centres. Provided data are comparable, we will then seek to pool small area data across centres, for analysis in a single

model, with setting and country included as fixed effects. Initially, null models will be conducted to determine the proportion of variance in incidence attributable to the neighbourhood level. We will then control for important individual level factors, including age and sex, and where possible, ethnicity, migration, social class or education. We will then test whether the models are improved (in both the within-centre and between-centre models) by standardised measures of neighbourhood exposures. Initially, this exposure set will be common across all centres to include (provided data are available - see above): population density, unemployment rate, overcrowding (people per household) and available measures of social fragmentation, including % single person households and residential turnover.

Two issues will receive detailed attention prior to the conduct of multilevel and spatial models.
First, the choice of small area will be given consideration, in order to ensure comparability across centres. This will partly depend on the availability of small area denominator data available in each centre, and the trade-off between statistical power and homogeneous neighbourhood units. Methods for combining smaller levels of geographies into larger areal units will be considered, including factor analysis and cluster analysis.
Second, Bayesian spatial models allow for the weighting of random effects to make different assumptions about the spatial dependency in incidence rates between neighbourhoods. Common model specifications include unstructured, structured, convoluted and spatial jump models, which make different assumptions about spatial dependency. Using local geographical and demographic knowledge, it may also be possible to determine a set of more realistic weights for the assessment of local neighbourhood dependency; that is, where a major road, river or other topographical feature acts as a physical barrier to social interactions between two adjacent neighbourhoods, it is possible to design a series of rules to weight spatial dependencies accordingly.

Other analyses/methods:

Involvement of external Parties (non EU-GEI):

External parties could be involved in each centre to obtain and/ or estimate some of the area variables and geographic details in each centre when these are not directly available in public statistics

IPR check:

Timeframe:

before June 2014 - based on information from the partners: decisions on the centres participating in the analysis and area definitions

before end 2014 - data provided to the analysis team

before May 2015 analyses and first draft of the article circulated to the group of authors

Submission of the manuscript: September 2015

Additional comments:

Because rural/urban and national differences will be used as additional (mainly confounder) variables [and because the working/writing group largely overlaps that on rural/urban and by centre analysis] we suggest that the analyses corresponding to this synopsis be done after the analysis of rural vs urban data.