Robustness and replicability of psychopathology networks

Network approaches to psychopathology hold that mental disorders arise from the interplay between symptoms in a network structure[1,2]. In the past few years, statistical techniques that estimate networks were developed and applied to many disorders[3]. As empirical findings start to accumulate, the question arising is which of these findings are robust and replicable. Here we evaluate the state of psychopathological network research based on three methodological criteria: model quality, precision, and replicability.

Model quality. One important quality of statistical modeling techniques is their ability to recover the "true" model that generated the data, a necessary prerequisite for justifying inferences based on network models. This is evaluated through: a) mathematical analysis: prove that a technique will recover the generating network model from data (e.g., by showing that it converges to the true model as sample size increases); and b) simulation studies: evaluate a technique's performance under various circumstances (e.g., for different network structures, sample sizes, and parameter settings).

Current state-of-the-art network techniques (i.e., pairwise Markov random fields[4]) have been vetted through mathematical proofs and simulation studies[5,6]: they efficiently recover the "true" model underlying the data. In general, such techniques minimize the false positive rate at the expense of statistical power. As a result, these techniques are more likely to omit "true" network connections, than to include spurious connections[5,6]. In sum, these techniques are vetted, conservative tools for estimating psychopathology network structures.

Precision and robustness. When a researcher has estimated a network from empirical data using vetted methodology, the question is to what extent the parameter estimates are precise: how robust are the results? For instance, if the relationship between self-worth and suicidal thoughts seems stronger than that between sleep difficulty and suicidal thoughts, it is necessary to investigate if model parameters are estimated with sufficient precision to justify this inference. If not, the result may not replicate in other samples.

Precision of network parameter estimates can vary considerably depending on factors such as sample size, network size, and network structure. Therefore, these factors must be assessed and reported on a case-by-case basis, by evaluating the statistical precision of parameter estimates (e.g., with confidence intervals) and the robustness of the model as a whole (e.g., investigating network structures in subsamples).

Dedicated freeware methodology for doing this recently became available[4], which allows researchers to report confidence intervals for estimated network parameters as an integral part of their results. This practice was quickly embraced by the majority of the network community, who now publish their work including detailed robustness checks. Naturally, results of such analyses should constrain the researcher's conclusions proportionately to their content: stronger claims (e.g., "insomnia is the most central node in the depression network") require stronger evidence than weaker claims (e.g., "insomnia is connected to the depression network").

Replicability. When network analysis seems to warrant an empirical conclusion (e.g., a particular symptom is highly central, or one network is more densely connected than another), the next question is whether the relevant phenomenon can be replicated in other samples. Ideally, replication research differs from the original study only in features that are deemed irrelevant to the phenomenon under investigation (e.g., by using a different sample from the same population). However, as is often the case in replication research, it is sometimes unclear whether differences between a study and its purported replication are relevant or not. For instance, if a network is first estimated on a community sample, and a replication is attempted in a patient sample, it may be unrealistic to assume that the same network holds in both populations. In such cases, studies probe not only a finding's replicability, but also its generalizability. Consequently, if inconsistent findings arise, this may either be because the phenomenon is unstable or illusory (i.e., the finding is not replicable) or because of substantively meaningful differences between studies (i.e., the finding is not generalizable to the context of the new study). In contrast, if an empirical phenomenon is observed consistently across studies, this provides compound evidence for both its replicability and generalizability[7].

Several recent empirical studies have evaluated the replicability of networks. The general picture which emerges is that network structures replicate and generalize well. For example, networks of major depression and generalized anxiety disorder symptoms are nearly identical in the US and Australia; post-traumatic stress disorder (PTSD) networks are similar across different populations and sources of trauma; and major depression networks are invariant across environmental and genetic risk factors (e.g., age of onset)[7,8].

Although network structures appear replicable and generalizable, detailed inferences based on them may be more susceptible to variation across studies. For example, the centrality of nodes seems to vary across PTSD networks, and a reported difference in network density between remitted and persistent major depression cases in adults was not fully replicated in an adolescent sample[8]. Future research should critically interrogate such findings to determine if inconsistency between studies is best characterized as a failure to replicate or a failure to generalize across contexts.

In conclusion, the model quality of network analysis techniques is good, while precision and robustness can now adequately be assessed with freely available methodological tools. Burgeoning replication research suggests that the structure of networks is typically consistent across studies, while stronger inferences based on these structures (e.g., centrality) have occasionally yielded mixed results.

Network analysis is a promising approach that may lead to significant improvements in research on and treatment of psychopathology[9], but researchers should be careful not to overstate causal conclusions based on network analysis as long as the causal interpretation of models has not been thoroughly investigated. The assessment of network robustness and replicability is an important step in this process and should be an important research focus in the next few years.

Denny Borsboom[1], Donald J. Robinaugh[1], The Psychosystems Group[1], Mijke Rhemtulla[2], Angélique O.J. Cramer[3]
[1]Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands; [2]Department of Psychology, University of California, Davis, CA, USA; [3]Department of Methodology and Statistics, Tilburg University, Tilburg, The Netherlands

The Psychosystems Group includes Sacha Epskamp, Eiko I. Fried, Lourens J. Waldorp, Claudia D. van Borkulo, and Han L.J. van der Maas.

1. Borsboom D. World Psychiatry 2017;16:5-13.
2. Cramer AOJ, van Borkulo CD, Giltay EJ et al. PLoS One 2016;11:e0167490.
3. Fried EI, van Borkulo CD, Cramer AOJ et al. Soc Psychiatry Psychiatr Epidemiol 2017;52:1-10.
4. Epskamp S, Borsboom D, Fried EI. Behav Res Methods 2018;50:195-212.
5. Bühlmann P, van de Geer S. Statistics for high-dimensional data: methods, theory and applications. Berlin: Springer, 2011.
6. Van Borkulo CD, Borsboom D, Epskamp S et al. Sci Rep 2014;4:5918.
7. Borsboom D, Fried EI, Epskamp S et al. J Abnorm Psychol 2017;126:989-99.
8. Fried EI. http://psych-networks.com/7-new-papers-network-replicability/.
9. McNally RJ. Behav Res Ther 2016;86:95-104.

DOI:10.1002/wps.20515

Accelerated biological aging in serious mental disorders

Individuals with serious mental disorders (SMDs) die at an earlier average age, even after controlling for suicide[1]. They are also at increased risk for developing somatic diseases that are typically associated with advanced age, such as cardiovascular diseases, metabolic syndrome, immune dysregulation and dementia[1].

The causes of this are likely multi-factorial, including genetic predisposition, biological changes set in motion by early life adversity, and lifestyle factors. Lifestyle factors, while obviously important, do not fully explain the increased mortality and morbidity in these individuals, and consequently, "accelerated biological aging" is increasingly being seen as an intrinsic factor in SMDs, at least in some individuals[2,3].

To the extent this hypothesis is true, the scope of pathophysiology in these illnesses would broaden considerably, and they might no longer be framed as only "mental disorders" or even brain diseases, but rather as whole-body, multi-system illnesses (or at least as illnesses with substantial somatic comorbidity), of which the psychiatric presentation is just the most readily observable pathology[3]. Understanding the mediators of such potential acceleration of aging should expand preventative and therapeutic opportunities to improve physical as well as mental health in affected individuals.

The notion of accelerated biological aging in SMDs is supported by reports of acceleration of certain biomarkers of age, such as leukocyte telomere length[2] and epigenetic age[4]. However, data on these biomarkers remain relatively sparse to this point, and several questions arise: a) Do these markers measure aging *per se*, or just the presence of factors that themselves mediate aging? b) Are these markers causally related to SMDs or just correlated with them? c) Is accelerated aging specific to particular psychiatric diagnoses or to certain physiological perturbations that traverse diagnostic boundaries? d) Do different aging biomarkers reflect the same or different underlying aging processes? Here I briefly review recent data pertinent to these questions.

Leukocyte telomere length and epigenetic age both significantly track chronological age, with correlation coefficients of −0.38 to −0.51 (for the former) and 0.96 (for the latter). Both of these markers significantly predict disease and mortality, strengthening the view that they are measurable markers of the aging process and of rates of aging. However, leukocyte telomere length and epigenetic age are independent predictors of chronological age and mortality risk[5]. Therefore, while they both measure processes that evolve with aging or are associated with aging, the specific processes are different, and their underlying mediators likely differ.

Telomere shortening can occur in response to inflammation, oxidative stress, stress hormones and other factors[2,3]. As such, it may signal the cumulative presence of a toxic cellular environment, rather than directly informing on the aging process itself. Indeed, leukocyte telomere length is often found to be inversely correlated with circulating inflammatory and oxidative stress factors[2,3]. Another major determinant of telomere shortening is a cell's mitotic history, since telomeres fail to fully replicate after each cell division, unless acted upon by the intracellular enzyme telomerase.

When cells reach a critically short telomere length, they may undergo replicative senescence, apoptosis, genomic instability or oncogenic transformation[2]. This can be especially problematic in tissues whose mitotic capacity is necessary for cellular replacement, such as hematopoietic stem cells and – of particular relevance to psychiatry – neuronal stem cells in the dentate gyrus of the hippocampus. Of great concern (and also of great preventative opportunity), early life stress, even *in utero*, has been associated with shortened leukocyte telomere length in newborns and in adults.

Telomeres in SMDs may progressively shorten with illness chronicity and/or severity, but, interestingly, even never-depressed girls at high genetic risk for developing depression already have short telomeres compared to girls at low genetic