

Community Reforms to Reduce Severity, Chronicity, and Disability of Psychosis: Do They Work?

In this issue of the *Journal*, Hegelstad et al. (1) report the results of the 10-year follow-up of the early Treatment and Intervention in Psychosis (TIPS) study in Norway and Denmark. The rationale for this study grew out of the observation that patients with a longer duration of untreated psychosis show a poorer symptomatic response to treatment, a more chronic course, and a poorer long-term outcome than those with briefer durations of untreated psychosis (2). The TIPS study was designed to enhance early detection and intervention of psychosis through a public education program that was implemented in two health care regions (referred to as early-detection sites). Subsequent cases of first-episode psychosis were detected and followed over time in the early-detection sites as well as in two other health care regions in which the educational program was not implemented (usual-detection sites). Reports on previous follow-ups showed reduced duration of untreated psychosis and less severe negative, depressive, and cognitive symptoms in patients from early-detection areas relative to those from usual-detection areas (3, 4). At the 10-year follow-up, although the groups no longer differed on these symptom dimensions, a higher percentage of patients from early-detection areas were characterized as “recovered” based on having achieved operationally defined standards of symptom remission and adequate levels of social and role functioning.

The TIPS project corresponds to a form of quasi-experimental design known as a community reform. Donald Campbell, a major figure in research methodology, is said to have affectionately referred to such designs as “queasy” experiments, in reference to the greater uncertainties inherent in a study context in which treatment assignment is not random and in which it is impossible to control for the myriad variables (dialects, ethnicities, religious affiliations, etc.) that may differ between the communities with and without reforms (5). Nevertheless, such designs are necessary to test any broad-based intervention that strives to alter health outcomes by changing community awareness and practices (6). Given that patients with psychotic symptoms enter treatment systems on average 1–2 years after initial onset (7), an intervention study designed to enhance early detection is precisely what is needed to determine whether reducing duration of untreated psychosis can affect the course and outcome of psychotic illness. The central issue here is whether, given several particularities related to the quasi-experimental nature of the design and the statistical approaches taken, it is safe to conclude that the TIPS community reform produced reliable benefits to patients in the long term.

The central issue here is whether, given several particularities related to the quasi-experimental nature of the design and the statistical approaches taken, it is safe to conclude that the TIPS community reform produced reliable benefits to patients in the long term.

Because a single rater interviewed all of the patients in the TIPS project, he could not be blind to which patients were from early- or usual-detection areas. Rater nonblindness poses a major problem given that the primary outcome variables (ratings of symptom severity and functioning) involve some intrinsic subjectivity and could be subtly influenced by the rater's knowledge of (and potential belief in) the study hypotheses. To address this issue, an independent rater evaluated a sample of videotaped interviews from the early-detection and usual-detection areas and was observed to be in good to excellent agreement (intraclass correlations between 0.6 and 0.85) with the original interviewer on the Positive and Negative Syndrome Scale and the Global Assessment of Functioning Scale. Nevertheless, as acknowledged by the authors, because of marked differences in dialects among sites, it was also not possible for the independent rater to be blind to the group status of the patients. This represents a flaw in the original study design, in that the experimental and control districts were not matched on factors such as dialect that could prevent awareness of raters to group assignments. Nevertheless, it is likely impossible to find two communities that are equivalent on all of the dimensions that could potentially compete with the intervention in explaining differences between the experimental and control regions (6).

Attrition is the bane of any longitudinal study, particularly when it may be non-random with respect to key study endpoints. In the TIPS study, attrition at the 10-year follow-up was higher in the usual-detection areas than in the early-detection areas, perhaps because patients in early-detection sites were easier to locate and access because of reduced barriers to participation in health care. Attrition at the 10-year follow-up was also selective, such that more severely impaired patients were overrepresented among those lost from the usual-detection group. Because the loss of more severe cases from usual-detection areas reduces, rather than increases, the likelihood of observing better outcomes among cases from early-detection sites, an attrition artifact cannot explain the finding of higher rates of clinical recovery in cases from early-detection sites, that is, the observed attrition bias only serves to strengthen, rather than diminish, confidence in the primary finding of the study. This selective attrition may, however, help to account for the loss of advantages in favor of early detection on several measures of symptom severity that had been observed in the 2- and 5-year follow-ups.

Patients from the early-detection group were more likely to have recovered clinically and to be working full time but were also less likely to be living independently and had higher excitability symptoms than patients from the usual-detection group. This apparently countervailing pattern raises the possibility that some or all of the differences may have resulted from chance. (The odds of observing two effects in the upper extreme are the same as those of observing two effects in the lower extreme in a random distribution.) Hegelstad et al. (1) note that when a family-wise Bonferroni correction of alpha (the threshold for statistical significance) is applied, the advantage in independent living for usual-detection patients does not survive while the other effects remain. This family-wise correction approach results in the use of different statistical thresholds for comparison of the primary outcome variable (recovery) for which no correction is applied ($\alpha=0.05$), ratings of symptoms and functioning (seven measures, $\alpha=0.05/7=0.0007$), and measures of work and independent living (three measures, $\alpha=0.05/3=0.017$). This procedure reserves the greatest statistical power for testing the effectiveness of the early-detection program on recovery as the primary endpoint, even though additional secondary endpoints are also evaluated. This procedure is generally ac-

cepted (8); however, one could argue that because the recovery measure is itself derived from seven separate measures of symptoms and functioning, the primary and secondary tests are nonindependent. If one were to adopt a single threshold that accounts for all of the comparisons conducted (11 measures; $\alpha=0.05/11=0.0045$), only the higher excitability symptoms among early-detection patients relative to usual-detection patients would remain, a pattern that would obviously not support the conclusion that early detection of first-episode psychosis leads to a less severe form of illness with better long-term outcome.

The TIPS study represents a unique and seminal contribution to the field of psychiatry, and the investigators are to be congratulated for conducting a very difficult long-term follow-up study. Nevertheless, depending on how you evaluate the impact of rater nonblindness, logical consistency of effects, and models of controlling for multiple comparisons, the advantages of early-detection programming may or may not remain.

References

1. Hegelstad WTV, Larsen TK, Auestad B, Evensen J, Haahr U, Joa I, Johannessen JO, Langeveld J, Melle I, Opjordsmoen S, Rossberg JJ, Rund BR, Simonsen E, Sundet K, Vaglum P, Friis S, McGlashan T: Long-term follow-up of the TIPS early detection in psychosis study: effects on 10-year outcome. *Am J Psychiatry* 2012; 169:374–380
2. Perkins DO, Gu H, Boteva K, Lieberman JA: Relationship between duration of untreated psychosis and outcome in first-episode schizophrenia: a critical review and meta-analysis. *Am J Psychiatry* 2005; 162:1785–1804
3. Larsen TK, McGlashan TH, Johannessen JO, Friis S, Guldberg C, Haahr U, Horneland M, Melle I, Moe LC, Opjordsmoen S, Simonsen E, Vaglum P: Shortened duration of untreated first episode of psychosis: changes in patient characteristics at treatment. *Am J Psychiatry* 2001; 158:1917–1919
4. Melle I, Larsen TK, Haahr U, Friis S, Johannessen JO, Opjordsmoen S, Simonsen E, Rund BR, Vaglum P, McGlashan T: Reducing the duration of untreated first-episode psychosis: effects on clinical presentation. *Arch Gen Psychiatry* 2004; 61:143–150
5. Trochim WMK: Donald T Campbell and research design. *Am J Evaluation* 1998; 19:407–409
6. Campbell DT: Factors relevant to the validity of experiments in social settings. *Psychol Bull* 1957; 54:297–312
7. Nordentoft M, Jeppesen P, Petersen L, Bertelsen M, Thorup A: The rationale for early intervention in schizophrenia and related disorders. *Early Interv Psychiatry* 2009; 3(suppl 1):S3–S7
8. Glimm E, Maurer W, Bretz F: Hierarchical testing of multiple endpoints in group-sequential trials. *Stat Med* 2010; 29:219–228

TYRONE D. CANNON, Ph.D.

From the Department of Psychology, University of California, Los Angeles. Address correspondence to Dr. Cannon (cannon@psych.ucla.edu). Editorial accepted for publication January 2012 (doi: 10.1176/appi.ajp.2012.12010094).

Dr. Cannon reports no financial relationships with commercial interests.