

## COMMENTARY

# What Type of Information Is Needed to Inform Mental Health Policy?

Roland Sturm\*

RAND, 1700 Main Street, Santa Monica, CA 90401, USA

### Abstract

The most valuable research integrates information from three levels of investigation: clinical efficacy, 'real life' effectiveness (including cost-effectiveness) and policy research. Successful applications of systematic reviews have largely been limited to clinical efficacy questions. The contribution of systematic reviews/meta-analyses to effectiveness and economic questions in mental health has been very minor and their contribution to inform policy is negligible. The latter is unlikely to change due to the different type of information that policy makers need.

As Gilbody and Petticrew argue in this issue,<sup>1</sup> systematic reviews can be useful decision-making tools, especially in the form of meta-analyses, which integrate the results of numerous individual studies quantitatively. Meta-analysis is undoubtedly a major methodological innovation. However, while systematic reviews/meta-analyses are featured in all leading medical journals, successful applications have largely remained limited to clinical efficacy questions. The contribution of systematic reviews/meta-analyses to economic questions has been very minor and their contribution to inform policy negligible—and I will argue that the latter is unlikely to change.

Health services researchers in the US commonly distinguish three levels of research. At the most basic level is information about clinical *efficacy*, i.e. whether or not a treatment works on carefully selected patients when applied in a controlled system. This information is the foundation for clinical decision making. At the next level is information about clinical and *cost-effectiveness*, i.e. how a treatment or organizational intervention, such as guideline implementation or quality improvement, affects typical patients in typical practice settings. This information is needed for decisions in larger organizations, such as health plans or hospital systems. At the highest macro level is information for *policy decisions* that affect organizational structures and incentives for providing health care through changes in health insurance, provider reimbursement and the legal system. Mental health policy in a decentralized pluralistic system like that in the US, where policy only affects health care indirectly by

changing the institutional framework, is concerned with this third level.

### Where Systematic Reviews Work: Treatment Efficacy and Clinical Decision Making

Systematic reviews have the most to contribute when there is a large number of studies that can be compared, which is most likely to be true for clinical efficacy trials. Protocols and research questions are similar; a main reason for the multitude of studies is that each research group only has access to a small number of patients. A large number of small studies with minor variations in patient selection and treatment protocols is the ideal situation for meta-analysis.

Guidelines on clinical practice, such as those from the Agency for Health Care Policy and Research (AHCPR) are often based on systematic reviews. The AHCPR depression treatment guidelines may be the most influential example in mental health based on a systematic review (although many other techniques are employed in the development of such consensus guidelines).<sup>2,3</sup> In addition, technology assessments, such as those conducted by the US Office of Technology Assessment, often include a systematic review of the literature on clinical efficacy as part of the assessment, similar to their UK counterparts discussed by Gilbody and Petticrew. Systematic reviews have already played an important role in this area and will continue to do so. All the successful examples cited by Gilbody and Petticrew fall in this area.<sup>1</sup>

### Where Systematic Reviews May Become Important: Real-World Effectiveness and Economic Evaluations of Treatment and Organizational Interventions

While clinical trials establish whether or not a treatment works in a controlled setting, it is also important to understand the *effectiveness* of a treatment, i.e. whether or not it also improves outcomes for typical patients treated in community practice settings.<sup>4–7</sup> While well known to readers of this journal, the distinction is not trivial and it is

\* Correspondence to: Roland Sturm, Ph.D. Senior Economist RAND, 1700 Main Street, Santa Monica, CA 90401, USA  
Source of funding: none declared

not always appreciated. Clinical trials provide treatments according to highly standardized or structured protocols, usually by trained study clinicians rather than usual care providers, and treatment costs are largely paid through research grants, changing provider and patient incentives. In contrast, a patient in actual practice settings may face varying costs across treatments, possibly resulting in differences in compliance (such as discontinuing an expensive drug) and consequently outcomes. In addition to different incentives, patients in efficacy studies are not representative of typical patients because clinical trials tend to select 'pure' clinical cases, exclude patients with comorbidities, and sample from the specialty sector in academic settings.

As Gilbody and Petticrew point out, the quality of primary clinical data is poor and little is known about 'real world' effectiveness of treatments, limiting the contribution of systematic reviews. An exception may be the extension of mental health efficacy research from tertiary settings to more representative primary care settings.<sup>8</sup> But the discussion of Gilbody and Petticrew also makes clear that meta-analysis cannot yet inform about the cost-effectiveness of antidepressants, which still relies on modeling, nor has it helped in providing precise summaries of the cost-effectiveness of assertive community treatment.<sup>1</sup> The lack of primary data is exacerbated when one considers the effectiveness of organizational interventions, which is a necessary step to change clinical practice, rather than of treatment interventions.

Most research funds continue to go towards clinical efficacy research, but the quality of primary data about the effectiveness of treatment and organizational interventions is improving. The National Institutes of Health in the United States have shifted emphasis from pure efficacy trials to effectiveness studies to increase the relevance for public health and public policy in some research areas. This has led to an increased focus on more broadly defined outcome measures, such as disability and health-related quality of life, not just clinical and disease-specific measures, such as blood pressure, remission status or psychiatric symptoms. Many effectiveness studies now also incorporate cost or cost-effectiveness evaluations of interventions compared to care as usual or standard treatments, and the National Institute of Mental Health explicitly recommends that intervention studies 'should include a cost-effectiveness component that uses the best methodology available'.<sup>9</sup> Kashner *et al.* discuss the development of the cost-effectiveness component of the Texas Medication Algorithm Project in this issue.<sup>10</sup> Other quality improvement studies have added similar cost components.<sup>5</sup>

Meta-analysis may become even more important as the lack of statistical precision for broader health and cost outcome measures is a ubiquitous problem. Contrasting a broad range of outcome and cost measures, we previously analyzed the implications for sample sizes and study design using data from prior mental health and substance abuse studies that spanned a wide range of practice settings and patient populations.<sup>11</sup> We concluded that while meaningful clinical symptomatic differences are often detectable with sample sizes of well under 100 per cell, detecting even

large changes in health-related quality of life generally requires several hundred observations per cell. Reasonable precision in cost estimates usually requires sample sizes in the thousands. Very few studies that incorporate quality of life or cost measures have such sample sizes, resulting in many (unreported) null findings and, due to publication biases favoring significant results, scientific publications that exaggerate true effects. Many effectiveness studies have additional features that challenge statistical power, such as clustered sampling designs, or randomization at the clinician or clinic, rather than patient, level.<sup>5,10</sup> Unfortunately, the shift towards broader health and cost outcome measures in effectiveness trials has not been accompanied by changes in study design that would allow researchers to address those questions with the statistical precision expected from traditional clinical trials.

To provide one example: The largest randomized trial of different antidepressant medications that explicitly was designed to analyze costs had 536 patients, divided into three cells.<sup>12</sup> While this may be an unusually large sample for a randomized trial, especially in a single institution, it does not have the statistical power to detect even large treatment cost differences. Health care costs in a 6 month period were slightly under \$2000 per patient in the fluoxetine group, about \$2100 in the imipramine group and over \$2300 in the desipramine group. However, based on the *t*-statistics reported in that paper, I calculated that a total sample of over 13 000 (4450 per cell) would have been needed to statistically identify a cost difference of 10% with 80% power. A 10% change on measures of quality of care may be considered a small intervention effect, yet for health care costs this could be a dramatic change. To put that into perspective, the Federal Mental Health Parity Act (a 1996 law in the United States that regulates the design of mental health benefits) and several state laws exempt employers if compliance would increase health care premiums by 1%.<sup>13-15</sup>

Short of extremely large multisite trials, which are so costly that very few could be supported, meta-analysis that combines data across multiple studies may be the most promising approach. A major strength of meta-analyses is to produce statistical power where individual studies do not. Thus, to the extent that they avoid the publication bias, meta-analyses may be needed to provide reliable insights about broader health and especially cost outcomes. In the area of quality improvement for depression care, the study design of several organizational interventions took the likely need for meta-analyses into account because the studies are individually too small for some questions, especially regarding economic outcomes and cost-effectiveness.<sup>5,16,17</sup> However, this group of studies may still be an exception and they are so new that the main results have yet to be published.

## Where Systematic Reviews Are Likely to Remain Limited: Health Care Markets and Policy

The US has a decentralized and pluralistic health care system, where health care policy means providing the

institutional framework, but not the direct provision of health care, which is left to market forces. Examples of health policy include legislation that regulates health insurance, employer mandates (such as the recent mental health parity debate in the US), liability (malpractice suits) or Medicaid reform. Data on treatment efficacy and effectiveness are important background material, but this is not the type of information policy makers need when they are considering legislation or have to implement regulations. Instead, they want data about the consequences of proposals, especially on the distributional consequences of proposals, i.e. who has to pay how much and who benefits.

This type of policy research remains sparse, for at least four reasons. The first reason is the relatively short window of opportunity to inform policy, often just a few months, leaving no time to put in place complex studies that require multiyear lead times. Second, the policy and health care environment changes much more quickly than human biology, and data collected 10 years ago may already be obsolete. Evidence from efficacy studies, in contrast, tends to be cumulative over many years. Third, multiple independent randomized trials (typical for clinical questions), which form the foundation for meta-analysis, are ill suited to answer policy questions. Policy often affects a large number of individuals indirectly and with relatively minor effects, requiring extremely large samples and randomization at the level of larger units (cities, counties, states), not randomization at the individual level. Finally, the funding infrastructure for independent policy research is much more limited than for clinical research. In short, the problem is the absence of data, as opposed to a wealth of contradictory primary data that systematic reviews can sort out. This is less the case for other areas of economic policy research, where the research base is much broader, for example welfare policy.<sup>18</sup>

The most urgent need in this third research area is for new data that track changes in markets and policies and provide a foundation for at least observational analyses.

One of the most prominent activities to fill this gap is the Robert Wood Johnson Foundation's Health Tracking Initiative, a multi-million-dollar initiative to track and report on changes in the US health care system. Regarding tracking changes in markets and policy for general health care, one major component is the Community Tracking Study (CTS), conducted by the Center for Studying Health System Change, which includes biennial longitudinal household surveys and complementary surveys of employers, physicians and insurance plans to track changes in health care delivery, insurance, access and costs of care.<sup>19</sup> Data from the first wave are already publicly available and form the foundation for a number of studies by other researchers to inform public and private health policy leaders about key issues in health care policy and market developments.

Mental health or drug abuse information lags behind, but a smaller part of the Health Tracking Initiative to collect data specifically on alcohol, drug and mental health issues was recently fielded.<sup>20</sup> The goal of this supplement is to quickly expand the data available to inform mental health policy debates by obtaining and linking information across

policy, service delivery and market levels. Data from the first wave will become publicly available by the end of 1999.

Summarizing my comments, I agree that systematic reviews can contribute to rational decision making, but primarily in areas that have an almost embarrassing richness of primary data. Clinical trials are the leading example. In contrast, mental health care policy in the United States suffers from a lack of data and research and is therefore not an area that could benefit much from increased emphasis on systematic reviews and meta-analyses.

## References

- Gilbody SM, Petticrew M. Rational decision-making in mental health: the role of systematic reviews. *J. Mental Health Policy Econ.* 1999; **2**: 99–106.
- Depression Guideline Panel. *Depression in primary care, volume 1. Detection and diagnosis.* US Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research, AHCPR Publication No. 93-0550. Rockville, MD: AHCPR, 1993.
- Depression Guideline Panel. *Depression in primary care: volume 2. Treatment of major depression.* US Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research, AHCPR Publication No. 93-0551. Rockville, MD: AHCPR, 1993.
- Wells KB. Treatment research at the crossroads: the scientific interface of clinical trials and effectiveness research. *Am. J. Psychiatry* 1999; **156**: 5–10.
- Wells KB. The Design of Partners in Care: Evaluating the Cost-Effectiveness of Improving Care for Depression. *Social Psychiatry Psychiatr. Epidemiol.* 1999; **34**: 20–29.
- Wells KB, Sturm R, Sherbourne CD, Meredith LS. *Caring for depression* Cambridge, MA: Harvard University Press, 1996.
- Simon G, Wagner E, VonKorff M. Cost-effectiveness comparisons using 'real world' randomized trials: the case of new antidepressant drugs. *J. Clin. Epidemiol.* 1995; **48**: 363–373.
- Schulberg HC, Katon W, Simon GE, Rush AJ. Treating major depression in primary care practice: an update of the Agency for Health Care Policy and Research Practice Guidelines. *Arch. Gen. Psychiatry* 1998; **55**(12): 1121–1127.
- National Advisory Mental Health Council. *Bridging science and service: A report by the National Advisory Mental Health Council's Clinical Treatment and Services Research Workgroup.* Bethesda, MD: National Institutes of Health, National Institute of Mental Health; 1998.
- Kashner TM, Rush AJ, Altshuler KZ. Measuring costs of guideline-driven mental health care: the Texas Medication Algorithm Project. *J. Mental Health Policy Econ.* 1999; **2**: 111–121.
- Sturm R, Unutzer J, Katon W. Effectiveness research and implications for study design: sample size and statistical power. *Gen. Hospital Psychiatry* 1999; **21**: 274–283.
- Simon GE, VonKorff M, Heiligenstein JH, Revicki DA, Grothaus L, Katon W, Wagner EH. Initial antidepressant choice in primary care: effectiveness and cost of fluoxetine vs tricyclic antidepressants. *J. Am. Med. Assoc.* 1996; **275**(24): 1897–1902.
- Sturm R, McCulloch J. Mental health and substance abuse benefits in carve-out plans and the Mental Health Parity Act of 1996. *J. Health Care Finance* 1998; **24**(3): 84–95.
- Frank RG, Koyanagi C, McGuire TG. The politics and economics of mental health parity laws. *Health Affairs* 1997; **16**: 108–119.
- Sturm R, Pacula RL. State mental health parity laws: cause or consequence of differences in use? *Health Affairs* 1999; **18**(5): 182–192.
- Meredith LS, Rubenstein LV, Rost K, Ford DE, Gordon N, Nutting P, Camp P, Wells KB. Treating depression in staff-model versus network-model managed care organizations. *J. Gen. Intern. Med.* 1999; **14**(1): 39–48.
- Rubenstein L, Jackson-Triche M, Unutzer J, Miranda J, Minnium K,

- Pearson M, Wells KB. *Design, dissemination and implementation of evidence-based interventions for depression in managed primary Care Practices*, WP 157, Center on Managed Care for Psychiatric Disorders, UCLA, Los Angeles, 1999.
18. Moffitt R. Incentive Effects of the US Welfare System: a Review, *J. Econ. Lit.* 1992; **30**: 1–61.
  19. Kemper P, Blumenthal D, Corrigan JM, Cunningham PJ, Felt SM, Grossman JM, *et al.* The design of the community tracking study: a longitudinal study of health system change and its effects on people, *Inquiry*, 1996; **33**: 195–206.
  20. Sturm R, Gresenz CR, Sherbourne, CD, Minnium K, Klap R, Bhattacharya J, *et al.* The design of health care for communities: a study of health care delivery for alcohol, drug abuse, and mental health conditions. *Inquiry* 1999; **36**: 221–233.