The Arts in Psychotherapy 37 (2010) 1–7

Contents lists available at ScienceDirect

The Arts in Psychotherapy

Applying levels of evidence to the psychiatric music therapy literature base

Michael J. Silverman, PhD, MT-BC*

University of Minnesota, School of Music, 100 Ferguson Hall, 2106 Fourth Street South, Minneapolis, MN 55455, United States

ARTICLE INFO

Keywords:
Levels of evidence
Music therapy
Mental health
Psychiatric
Evidence-based practice

ABSTRACT

In an era of evidence-based practice (EBP), it is becoming increasingly important to distinguish the quality of research studies and synthesize results so they can be applied to clinical practice. Thus, in an attempt to categorize research and amalgamate results, scholars have developed various hierarchical levels of evidence to differentiate research implications. However, these levels of evidence have not yet been applied to the psychiatric music therapy literature base. The purpose of this paper was to discuss and identify the levels of evidence and apply well-established levels of evidence to the psychiatric music therapy literature base. Results indicated a lack of randomized controlled trials and overall low level of evidence. Further, regardless of taxonomy applied, most studies met criteria for the lowest level of evidence. This finding is congruent with the levels of evidence of other well-established psychosocial treatments for psychiatric consumers. Limitations, generalizations, and implications for research and clinical practice are provided.

Introduction and review of literature

In today's continuously evolving healthcare system, the identification, synthesis, and application of research evidence is becoming increasingly important. For clinicians to design interventions based upon the best available evidence, it is imperative to differentiate and understand the levels of research evidence (Fineout-Overholt, Hofstetter, Shell, & Johnson, 2005; Fisher & Wood, 2007). Psychiatric mental healthcare is no exception and requires the utilization of the best current research available (Rice, 2008). However, researchers in psychiatric mental healthcare have repeatedly articulated the difficult nature of this task due to the subjective experience of the consumers (Salmond, 2007) and the limited use of clinical trials (March et al., 2005).

It has been continually articulated that research is the best way to develop an understanding of the effects of various treatments, consumer responses to those treatments, and the resultants. As a result of this process, research evidence has become the basis for clinical healthcare decisions (Stevens, 2005). Evidence-based medicine (EBM) was developed in an attempt to better provide quality outcomes that were based upon existing research (Elstein, 2004; Sackett, Rosenberg, Gray, Haynes, & Richardson, 1996). Evidence-based practice (EBP) was a resultant of this movement toward clinical practice based upon systematic investigation and consumer preference. The term "evidence-based" (EB) was first used in 1990 (Eddy, 1990, 2005) while the term EBM first appeared in the medical literature in 1992 (Guyatt et al., 1992). David Eddy, who applied mathematics to practitioner decision making, was the first to use EBM as a term (Arzin & Goldman, 2005). The EB movement was founded by British epidemiologist Dr. Archie Cochrane (Melnyk & Fineout-Overholt, 2005), often referred to as the "godfather of evidence-based medicine" (Arzin & Goldman, p. 70) and strong supporter of the randomized controlled clinical trial. While a thorough discussion of EBP and its history is certainly beyond the scope of this paper, there is a need to articulate various levels that comprise EBP. With the exception of Edwards (2005), these hierarchies have been discussed in other fields, but not specifically in psychiatric music therapy, there is a pressing need to articulate the various levels of EBP in psychiatric music therapy in order for the research base to continue to develop. Therefore, the purpose of this paper is to review the various levels of evidence and apply these to the existing psychiatric music therapy literature base. A brief introduction and discussion of levels of evidence follows.

Systematic reviews and meta-analyses are typically considered the top level of scientific evidence (Level I). In an era when academics and practitioners are bombarded with a seeming overabundance of studies, this level of evidence is also most efficient: the health literature research base is expanding at a rapid rate and is expected to double in 19 years (Benner & Leonard, 2005). Thus, systematic reviews can be a convenient and time-efficient method to assess the literature base. Scholars have noted that these reviews should also include unpublished studies to reduce publication bias (Anderson & Beck, 2003; Egger, Smith, & Sterne, 2001). Additionally, when available, many scholars have recommended that only randomized controlled trials (RCTs) should be included in systematic reviews and meta-analyses. However, this is often not possible,

* Tel.: +1 612 624 1091.
E-mail address: silvermj@umn.edu.

0197-4556/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.aip.2009.11.005

as in the case of many psychosocial treatment studies involving mental health consumers.

RCTs are typically considered the best evidence with the exception of systematic review/meta-analysis. They are often – but not always – considered Level II Evidence, pending upon the specific hierarchy utilized. However, other types of research comparing the presence of the group with the illness (or phenomenon of interest) with a comparison group can sometimes be considered Level II Evidence (Rice, 2008). Some researchers have noted that this level can include clinical trials that are not randomized and cohort studies (Fisher & Wood, 2007). Although Cochrane (1972) advocated for the use of RCTs and they are still largely considered the most rigorous type of evidence, there remains a great deal of controversy regarding the superiority of this design (Baldwin, 2006; Goldsmith, Bankhead, & Austoker, 2007; Joseph, 2008) and, more generally, what constitutes Level II Evidence.

Though well-designed and controlled meta-analyses are important in allocating funding and shaping clinical practice, quasi-experimental studies should not be discounted as they may constitute the best available evidence in light of the numerous complications that researchers encounter (Drake et al., 2001). While open clinical trials lacking comparison groups are generally considered to lack vigor due to lack of randomization and potential for error (Drake et al., 2001), they may represent the best available evidence. Additionally, when no other evidence exists, qualitative and descriptive studies can be used to provide initial, non-RCT data concerning interventions and how they may be potentially studied at other levels. However, although these types of studies are considered a lower level of evidence, it may be the only available evidence available, especially in rare conditions or illnesses that are particularly difficult to systematically investigate (Fisher & Wood, 2007). Furthermore, it is no longer considered sufficient to use traditions and opinions in today's EBP society (Rice, 2008). Regardless of taxonomy, these are considered the lowest level of evidence. While these opinions may be convenient in that they permit the development of guidelines when no research exists, they can often reflect biases rather than effectiveness.

Although qualitative studies can be very useful in describing perspectives and experiences and there are scientific methods for determine overall effects, the systems for appraising qualitative research are still in their infancy (Harden et al., 2004). As much of the nursing evidence has utilized a qualitative approach to better understand the unique experiences of patients, Salmond (2007) developed a pyramid of descriptive experiences to summarize qualitative research. In descending order from the top of her pyramid (with the top of the pyramid representing the highest level), Salmond categorized a qualitative evidence hierarchy: (1) systematic review of descriptive and qualitative studies, (2) evidence from a single descriptive or qualitative study, (3) expert opinion, expert committees, and (4) evidence based on quantitative approaches.

Specific levels of evidence

Experts in EBP have developed a number of taxonomies or hierarchies articulating the strength – or lack thereof – of quantitative evidence (Rice, 2008; Stevens, 2005). This ranking is considered a basic principle of EBP (Fineout-Overholt et al., 2005). The taxonomies are typically organized around various research designs. Concerning the various research designs, the Agency for Healthcare Research and Quality (Agency for Healthcare Research and Quality, 2002; AHRQ, 2002) also articulated three domains from which grades of research strength are based: quality, quantity, and consistency (Crowther & Cook, 2007; West et al., 2002).

The U.S. Preventative Task Force (2003) developed a system that is based on a straightforward ABC grading system. On the other end of the spectrum is the "Oxford Center" classification, which is both global and complex (Phillips et al., 2001). The AHRQ also identified systems for appraising evidence (AHRQ, 2002). This report was exhaustive and identified the diversity of protocols for the evaluation of scientific evidence: twenty systems were found that appraised systematic reviews, 49 systems were found that appraised RCTs, 19 systems were found that appraised observational studies, 18 systems were found that appraised diagnostic tests, and 40 systems were found that appraised the strength and limitations of a body of evidence (West et al., 2002). The scholars authoring this report did not recommend that one system should be used over another but did articulate the lack of standardization for rating systems (AHRQ, 2002). Additionally, the National Institute for Health and Clinical Excellence (NICE) in the United Kingdom provides very specific sets of guidelines (www.nice.org.uk).

In a report of the 2005 Presidential Task Force on EBP, Levant (2005) did not include an ordered/numbered hierarchy but did differentiate between various levels of research. These levels are described in Table 1.

Similar to Levant (2005), Devereaux and Yusuf (2003) composed an organized hierarchy but did not assign levels. Their descriptions are depicted in Table 2.

Rice (2008) integrated a number of approaches in a hierarchy. In this article, Rice combined the U.S. Preventative Services Task Force clinical grading with data levels and quantitative research designs. Rice's integration of rankings is depicted in Table 3.

Numerous researchers have attempted to apply levels of evidence to the nursing literature. Due to the need to incorporate different types and levels of studies, these taxonomies have been more broad (Fineout-Overholt & Johnston, 2006), and thus potentially applicable for a number of professions. Melnyk and

Table 1
Descriptions of evidence (Levant, 2005, pp. 7–8).

Clinical observation (including individual case studies) and basic psychological science are valuable sources of innovations and hypotheses (the context of scientific discovery).
Qualitative research can be used to describe the subjective lived experience of people, including participants in psychotherapy.
Systematic case studies are particularly useful when aggregated as in the form of practice research networks for comparing individual patients to others with similar characteristics.
Single-case experimental designs are particularly useful for establishing causal relationships in the context of an individual.
Public health and ethnographic research are especially useful for tracking the availability, utilization, and acceptance of mental health treatments as well as suggesting ways of altering them to maximize their utility in a given social context.
Process-outcome studies are especially valuable for identifying mechanisms of change.
Studies of interventions as delivered in naturalistic settings (effectiveness research) are well suited for assessing the ecological validity of treatments.
Randomized clinical trials and their logical equivalents (efficacy research) are the standard for drawing causal inferences about the effects of interventions (context of scientific verification).
Meta-analysis is a systematic means to synthesize results from multiple studies, test hypotheses, and quantitatively estimate the size of effects.

Table 2
Hierarchy of evidence for evaluating a study (Devereaux & Yusuf, 2003).

Systematic review of several large randomized controlled trials
Single large randomized controlled trial
Systematic review of several small randomized controlled trials
Single small randomized controlled trial
Systematic review of several cohort studies
Single cohort study
Systematic review of several case–control studies
Single-case–control study
Systematic review of several cross-sectional studies
Single cross-sectional study
Case series