

一所高校心理健康教育中心的心理咨询效果 ——基于实践的证据

闫玉朋¹ 江光荣¹ 应梦婷² 于丽霞^{1,3} 林秀彬¹ 李丹阳¹

1. 华中师范大学心理学院, 湖北省人的发展与心理健康重点实验室, 青少年网络心理与行为教育部重点实验室, 武汉;
2. 武汉体育学院健康科学学院心理学系, 武汉;
3. 华中师范大学大学生心理健康教育中心, 武汉

摘要 | 目的: 考察某高校的心理咨询效果。方法: 数据来自于十年前国内一所高校心理健康教育中心使用中文版心理咨询效果问卷(OQ-45.2)从第二次会谈开始持续收集的效果数据, 为时一年半, 期间测量了103名当事人随着咨询进程发生的改变, 剔除无效数据后, 共计63名当事人, 其中临床样本包括31名当事人。结果: 总样本前后测效应量为中等效应量($d=0.57$); 临床样本前后测效应量为大效应量($d=1.22$), 优于基准; 总样本康复率22.2%, 改善率14.3%, 无变化率57.1%, 恶化率6.3%, 趋势与基准一致, 无显著差异; 将总样本分为康复与未康复两类后, 总样本康复率优于基准($p<0.05$)。结论: 本研究提示该中心可能取得了良好的咨询效果。

关键词 | 高校心理健康教育中心; 心理咨询效果; 基于实践的证据; 基准法

Copyright © 2021 by author (s) and SciScan Publishing Limited

This article is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/). <https://creativecommons.org/licenses/by-nc/4.0/>



效果历来是心理咨询领域的一个重要研究主题, 随机临床试验(random clinical trial, RCT)因其严格的实验操控和高内部效度而成为检验心理咨询效果的金标准。多年来, 大量RCT及基于RCT的元分析为心理咨询提供了有力的效果证据, 从而巩固了心理咨询的合法地位。然而, RCT的低生态效度让人质疑其能否代表自然设置中的心理咨询的效果^[1, 2]。

于是, 自然设置下的咨询效果考察也就势在必行了。这一类研究属于实践取向的研究(practice-focused research), 也被称为基于实践的研究(practice-based study)或基于实践的证据(practice-based evidence)^[3, 4]。基于实践的研究证明了多个疗法在常规实践中的效果, 比如以人为中心疗法^[5]、认知行为疗法^[6, 7]和

通讯作者: 江光荣, 华中师范大学心理学院教授、博导, 研究方向: 临床与咨询心理学、学校心理辅导与心理健康教育。

文章引用: 闫玉朋, 江光荣, 应梦婷, 等. 一所高校心理健康教育中心的心理咨询效果——基于实践的证据[J]. 心理咨询理论与实践, 2021, 3(4): 206-213.

<https://doi.org/10.35534/tppc.0304026>

心理动力—人际疗法^[8]。Stiles 等人^[9]甚至同时考察了以人为中心疗法、认知行为疗法和心理动力疗法,验证了元分析所发现的渡渡鸟效应(不同疗法效果等值)。而且,一些设置(如大学咨询中心、员工援助项目全国健康维护组织、大学心理咨询训练门诊和州心理健康服务机构等)^[10, 11]的效果也得到检验。此外,基于实践的研究可进一步探究咨询师效应(咨询师对咨询效果的作用)、咨询师的专业性发展及咨询机构效应(咨询机构对咨询效果的作用)当事人的改变模式、剂量—效果关系(会谈次数与效果的关系)、预期治疗反应、正式反馈等问题^[4, 12, 13]。

在国内,我们目前仅发现两个相对规范的基于实践的效果研究。秦旻^[14]采用自编的门诊心理治疗效果问卷,在三家三级甲等综合医院的心理门诊,考察接受两次及以上心理咨询且自愿参与研究的55名当事人的咨询效果(平均会谈次数 5.47 ± 3.71),仅以可信改变指数^[15]为指标,发现18名(32.7%)当事人在咨询结束后达到了可信的临床改善。李钰静^[16]在一家医院的心理门诊、三所私立心理咨询室和三所高校心理咨询中心,采用中文版的心理咨询效果问卷(outcome questionnaire-45.2, OQ-45.2),收集到41名接受两次及以上咨询(平均会谈次数 3 ± 2)且初始测量达到临床临界分的当事人的数据,以Jacobson和Truax^[15]的临床显著性改变(包括分数是否跨越临床划界分与改变分是否达到可信改变指数)为指标,发现8人(19.5%)康复,8人(19.5%)改善,24人(58.6%)无改变,1人(2.4%)恶化。尽管这两个研究样本量小,也未说明样本对所调查机构当事人总体的代表性,但对于国内的基于实践的证据来说,无疑具有开创意义。

几乎与上述两个研究在同一时间,湖北省某高校建立了心理咨询过程—效果研究数据库。该数据库包括从2009年5月到2011年1月期间该高校心理健康教育中心(以下简称“该中心”)103名当事人的心理咨询过程数据、效果数据、录音及录像。其中部分数据曾用于考察一系列重要的过程变量和改变机制。在重新整理该数据库的过程中,我们发现,该数据库所得到的效果数据在一定程度上可代表该中心的实践情况,可作为国内又一个基于实践的效果证据。

1 对象与方法

1.1 对象

所有研究对象均为前来该中心咨询的当事人。该中心也是湖北省青少年心理健康教育中心,主要为本校大学生和省内青少年提供个体心理咨询,通常限制为每学期8次会谈,但允许咨询师和当事人灵活调整。同时,该中心为本校临床与咨询心理学专业方向的硕士生和博士生的实习基地。对于不适合接受心理咨询或需要其他辅助的当事人(比如疑似精神分裂症患者、发作期的双相情感障碍患者、严重自杀风险的当事人、中重度抑郁症等),咨询师会转介当事人到医院的精神科门诊接受评估或治疗,或是进入该机构的危机干预程序,结合学校、家庭、医院的多方面力量共同工作。部分经医生评估后被推荐同时接受药物治疗和心理咨询的当事人,也纳入本研究之中。

103名当事人参与了本研究,共进行了520次会谈,平均会谈次数 5.05 ± 3.81 ,比较接近2010年该中心平均会谈次数(4.97),或可说明该样本对该中心当事人总体具有一定的代表性。根据档案数据推断,参与该数据库建设的当事人约占同时间中心接待当事人总体的15.5%。初次会谈前,研究者向当事人介绍该研究,签署知情同意书,收集当事人基本信息及其他一些变量,从第二次会谈前开始进行效果测量。故35名(34.0%)仅进行一次或两次会谈的当事人数据因无法计算前后测改变量而被删除。另有

5名(4.85%)当事人因首次或末次效果测量缺失而被移除。最终共63名当事人纳入本次分析。

在63名当事人中,男11名(17.5%),女50名(79.4%),缺失2名;年龄在18到28之间(平均年龄 21.76 ± 2.28 , $n=55$);大一、大三均为10人(各15.9%),大二14人(22.2%),大四6人(9.5%),研究生22人(34.9%),缺失1人;21人(33.3%)此前有咨询经历。会谈次数 7.05 ± 3.53 (范围3~18)。根据李钰静和骆宏^[17]制定的中文版OQ-45.2的临床临界分(63分),即,当事人初始分 ≥ 63 则为心理健康状态异常(临床样本),共有31名当事人在初次效果测量时 ≥ 63 分,是为临床样本。

38名咨询师(7男,31女,专职2人,兼职14人,博士实习生6人,硕士实习生16人)为这63名当事人提供心理咨询。由于咨询年限会随时间有轻微变化,咨询师在不同当事人处所填的理论取向也不尽相同,故二者按人次计算。咨询年限为 3.95 ± 3.69 ($n=59$);精神分析/心理动力取向 2.77 ± 0.95 ,人本/以人为中心取向 4.15 ± 0.67 ,认知/认知行为取向 3.18 ± 0.76 (范围1~5,5分最高, $n=62$)。

1.2 工具

中文版OQ-45.2。英文原版OQ-45.2是使用最为广泛的心理咨询效果测量工具之一^[18],共45道自陈题目,包括症状困扰、人际关系和社会角色功能三个维度,分数越高,表明当事人越严重。李钰静和骆宏^[17]获得原作者授权修订的中文版OQ-45.2,也是45道自陈题目,内部一致性信度0.91~0.92,重测信度0.71。本研究所采用的OQ-45.2为建设数据库时获原作者授权由英文原版翻译而来^[19],与李钰静和骆宏^[17]修订的题目在语义上没有区别。本研究中OQ-45.2第二次会谈的内部一致性信度为0.92,最后一次会谈的内部一致性信度为0.93。

1.3 基准选择

基准,即可用于同类事物比较的标准。基于实践的研究往往选择在某些方面(如样本、所使用的疗法、效果测量根据)有可比性的RCT、元分析或大样本的基于实践的证据作为基准^[3,4],通过和基准对比,形成对其结果的更为清晰的定位。本研究最终选取了国外两所高校心理健康中心使用原版OQ-45.2测量的效果数据作为基准。其中,Minami等^[11]仅计算了临床样本的效应量,故将其作为基准与本研究的临床样本进行对比。他们分析了美国犹他大学心理咨询中心8年数据,选取了会谈次数在两次及以上且初始效果测量超过临床临界分(英文版OQ-45.2的临床临界分为64)的2672个个案,占该中心同时间接待总量的34.93%;纳入该数据分析的咨询师共148名,以女性为主,理论取向多为整合取向。该咨询中心也以服务大学生为主,且接纳多个水平、多种心理健康专业的实习生。Hansen等^[10]则报告了其总样本的临床显著改变状态,故将其作为基准与本研究的总样本进行对比。他们考察的是美国西部一所大学心理咨询中心1188名当事人的数据,删除了因一次结束咨询无法计算改变量的当事人数据(占总样本约33%),其他信息不详。

1.4 统计方法

使用SPSS23.0进行初步分析,计算总样本和临床样本的基线分数和咨询后分数。

使用GraphPad Prism 8.3.0进行统计分析。采用Welch-corrected t 检验比较总样本和临床样本与其基准会谈次数的差异;使用独立样本 t 检验比较临床样本与其基准基线分数及咨询后分数的差异;采用 χ^2 检验比

较总样本与其基准临床显著改变状态的差异；采用 Fisher 的精确检验比较总样本与其基准康复率的差异。

计算咨询效果的效应量。不同于常见的以联合标准差作为分母计算两组差异效应量，心理咨询效果的前后测差异效应量多以初始分数为分母，因为它不受反复测量和咨询的影响^[11]。

计算临床显著性。李钰静和骆宏^[17]根据 Jacobson 和 Truax^[15]计算临床显著性的方法，制定了中文版 OQ-45.2 的临床临界分（63 分）和可信改变指数（17 分）。根据这两个指标可对当事人的临床显著改变状态进行划分。具体来说，当事人基线分 ≥ 63 ，咨询结束时分数 < 63 ，且降低 ≥ 17 分，是为康复；当事人分数变化未跨越临界分，但降低 ≥ 17 分，是为改善；当事人改变分 < 17 分，是为无改变；当事人增加 ≥ 17 分，是为恶化。本文据此对当事人的临床显著改变状态进行分类和计数。

2 结果

2.1 会谈次数及其对比

各样本的平均会谈次数见表 1。Welch-corrected t 检验结果表明，总样本的会谈次数高于 Hansen 等^[10]，本研究临床样本与 Minami 等^[11]的平均会谈次数无显著差异。

表 1 各样本平均会谈次数的比较 ($\bar{N}_{se} \pm s$)

Table 1 Means, standard deviations, and comparisons of the number of sessions

本研究总样本 ($n=63$)	Hansen 等 ^[10] ($n=1188$)	t 值	本研究临床样本 ($n=31$)	Minami 等 ^[11] ($n=2672$)	t 值
7.05 ± 3.53	5.80 ± 5.40	2.65**	7.45 ± 3.71	6.84 ± 8.72	0.89

注：** $p < 0.01$ 。

2.2 效应量及其对比

该中心临床样本与其基准 Minami 等^[11]的基线分数和咨询后分数见表 2。独立样本 t 检验结果表明，该中心临床样本基线平均分低于其基准，咨询后平均分也低于其基准。总样本的基线平均分为 62.86 ± 21.01 ，咨询后平均分为 50.84 ± 20.4 ，其基准 Hansen 等^[10]未提供基线分数和咨询后分数，故总样本无法与之比较。

表 2 临床样本与其基准的基线分数和咨询后分数的比较 ($\bar{x} \pm s$)

Table 2 Means, standard deviations, and comparisons of outcomes at baseline and post-termination

	本研究临床样本 ($n=31$)	Minami 等 ^[11] ($n=2672$)	t 值
基线	79.48 ± 14.73	87.44 ± 16.44	2.68**
咨询后	61.55 ± 20.61	71.39 ± 22.16	2.46*

注：* $p < 0.05$ ，** $p < 0.01$ 。

除 Hansen 等^[10]未提供效应量外，其余各样本的效应量、效应量平均改变量和平均改变分数见表 3。就总样本而言，该中心的咨询效果为中等效应量；就临床样本而言，该中心得到的效应量为大效应量，

且明显大于 Minami 等^[11]报告的效应量。平均改变的效应量和平均改变分数比较接近。

表 3 各样本的效应量、平均改变的效应量和平均改变分数

Table 3 Effect size estimates, effect size estimates divided by the number of sessions, and means of outcome changes

研究	效应量 d	d/N_{se}	$\bar{x}_{基} - \bar{x}_{后} / N_{se}$
本研究总样本	0.57	0.08	1.70
本研究临床样本	1.22	0.16	2.41
Minami 等 ^[11]	0.98	0.14	2.35

2.3 临床显著改变状态极其对比

除 Minami 等^[11]未提供临床显著改变状态的数据，其余样本的临床显著改变状态见表 4。本研究总样本的临床显著改变状态与其基准 Hansen 等^[10]整体趋势较为一致：两组“康复”与“改善”的比例之和均在 30% 到 40% 之间；“无改变”的比例均为最高，且十分接近；“恶化”的比例都较低。 χ^2 检验表明两组在临床显著改变状态上无显著差异（ $\chi^2=5.93$ ， $p>0.05$ ）。考虑到两组康复率看起来差异较大，特将两个样本划分为康复与未康复（包括改善、无改变和恶化）两类，Fisher 的精确检验表明本研究总样本的康复率显著高于 Hansen 等^[10]（ $p=0.03<0.05$ ）。

表 4 临床显著改变状态

Table 4 Classification of outcome groups based on client clinically significant change status

研究	临床显著改变状态人数（百分比）			
	康复	改善	无改变	恶化
本研究总样本	14 (22.2%)	9 (14.3%)	36 (57.1%)	4 (6.3%)
Hansen 等 ^[10]	150 (12.6%)	239 (20.1%)	684 (57.6%)	115 (9.7%)
本研究临床样本	14 (45.2%)	2 (6.5%)	13 (41.9%)	2 (6.5%)

3 讨论

总的来说，该中心的心理咨询有良好的效果。从效应量看，临床样本效果优于基准，改变速度与基准相当；从临床显著改变状态看，总样本整体状态与基准一致，且康复率显著更高。但在理解这一结果时，需要注意几个问题：一是初始严重程度和咨询次数的影响。初始严重程度与效果关系密切，一些同样使用 OQ-45.2 的研究发现当事人初始状态越严重，改善越大^[20]。而本研究临床样本的初始严重程度低于其基准 Minami 等^[11]，更加证明本研究临床样本的效果。本研究总样本咨询次数大于其基准，更高的康复率可能受到咨询次数的影响。二是基于实践的研究常常难以避免的局限的影响^[7, 9, 21, 22]，包括：（1）没有对照组，不能确定性地将所有改变仅归于咨询。但考虑到多数当事人在觉察到自身问题很久之后才寻求心理咨询^[23]，将咨询效果多数归于咨询也有一定的合理性；（2）由于数据缺失或样本代表性而造成对结果的影响。虽然从平均会谈次数看，本研究的样本具有一定代表性，但仍不能对其代表性进行确切估计，也就难以估计其对结果的影响；（3）未收集当事人的服药信息，无法排除药物所带来的效果；（4）未用手册、

持续的训练、系统督导及遵循度评估等保证所进行的治疗是合格的或具有保真度（integrity）的治疗。这也正是基于实践的研究的优势所在——高生态效度。三是数据收集时间的影响。本研究第二次会谈开始收集收据，一方面，这可能缩小了改变分，从而略微降低了效应量；另一方面这也造成两次结束的当事人无法纳入本次分析，可能又增加了效应量。这些局限一定程度影响到该数据的可推广性，限制了对其进一步分析、解释的空间，但考虑基于实践的证据较强的生态效度，以及当前国内此类研究的严重缺乏，对该数据的分析，仍能够为国内心理咨询提供一个基于实践的证据，并为未来的研究提供启发。

另一个值得关注的问题是，本研究与国内外的研究的发现^[11,16]一致，有超过一半的当事人“未改变”，或者说“未发生可信的改变”。本研究临床样本有超过40%的当事人未改变，这意味着有相当比例的较为严重的当事人，离开咨询时未发生实质性改变，这有可能是因为当前的咨询存在较大局限，仅适用于特殊的当事人群体，也有可能是因为某个或某些当事人特征对效果有重要影响，这些都有待进一步研究。

由于本研究样本较小，且代表性存在一定局限，故未进一步考察效果的影响因素，而当前得到的多数结果也都是尝试性的，亟待在国内的常规实践中开展系统的、大规模的基于实践的研究，从而进行检验。这不仅能提供更有力的效果证据，同时也可推动进一步的研究与实践。具体来说，需要做的工作可能包括以下几个方面。首先，将系统、可靠的效果评估作为咨询机构日常实践的常规工作。一方面，监控效果可为机构决策和研究提供基础；另一方面，可将效果评估反馈给咨询师，能够提升咨询效果^[24,25]。其次，完善数据收集过程。从当事人预约时即开始收集效果数据，以建立起自然设置中的等待对照组，从而获得对咨询效果更为合理的估计；收集当事人的诊断信息，从而可与相关RCT进行对比；收集当事人的服药信息，从而可对比药物治疗和咨询的作用；由咨询师对当事人的结案情况进行登记，可分为达成目标的结案、未达成目标的结案、脱落等；在咨询结束后定期追踪当事人的效果。再次，进一步探索无改变、脱落及恶化等特殊的当事人群体。最后，建立多机构的实践与研究合作网络^[4]，致力于研究者和实践者共同研究真正有实践意义的问题，从而进一步改善实践。

参考文献

- [1] Weisz J R, Weiss B, Donenberg G R. The lab versus the clinic: Effects of child and adolescent psychotherapy [J]. *American Psychologist*, 1992, 47 (12): 1578-1585.
- [2] Westen D, Novotny C M, Thompson-Brenner H. The empirical status of empirically supported psychotherapies: Assumptions, findings, and reporting in controlled clinical trials [J]. *Psychological Bulletin*, 2004, 130 (4): 631-663.
- [3] Barkham M, Stiles W B, Lambert M J, et al. Building a rigorous and relevant knowledge-base for the psychological therapies. In: Barkham M, Hardy GE, Mellor-Clark J. *Developing and Delivering Practice-based Evidence: A Guide for the Psychological Therapies* [M]. Chichester, United Kingdom: Wiley, 2010: 21-61.
- [4] Castonguay L G, Barkham M, Lutz W, et al. Practice-oriented research: Approaches and applications. In Lambert MJ. *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change, Sixth Edition* [M]. Hoboken, New Jersey: Wiley, 2013. 85-133.
- [5] Gibbard I, Hanley T. A five-year evaluation of the effectiveness of person-centred counselling in routine clinical practice in primary care [J]. *Counselling and Psychotherapy Research*, 2008, 8 (4): 215-222.
- [6] Persons J B, Bostrom A, Bertagnolli A. Results of randomized controlled trials of cognitive therapy for

- depression generalize to private practice [J]. *Cognitive Therapy and Research*, 199, 23 (5): 535-548.
- [7] Westbrook D, Kirk J. The clinical effectiveness of cognitive behaviour therapy: Outcome for a large sample of adults treated in routine practice [J]. *Behaviour Research and Therapy*, 2005, 43 (10): 1243-1261.
- [8] Paley G, Cahill J, Barkham M, et al. The effectiveness of psychodynamic-interpersonal therapy (PIT) in routine clinical practice: A benchmarking comparison [J]. *Psychology and Psychotherapy: Theory, Research and Practice*, 2008, 81 (2): 157-175.
- [9] Stiles W B, Barkham M, Mellor-Clark J, et al. Effectiveness of cognitive-behavioural, person-centred, and psychodynamic therapies in UK primary-care routine practice: Replication in a larger sample [J]. *Psychological Medicine*, 2008, 38 (5): 677-688.
- [10] Hansen N B, Lambert M J, Forman E M. The psychotherapy dose-response effect and its implications for treatment delivery services [J]. *Clinical Psychology: Science and Practice*, 2002, 9 (3): 329-343.
- [11] Minami T, Davies D R, Tierney S C, et al. Preliminary evidence on the effectiveness of psychological treatments delivered at a university counseling center [J]. *Journal of Counseling Psychology*, 2009, 56 (2): 309-320.
- [12] Firth N, Saxon D, Stiles W B, et al. Therapist and clinic effects in psychotherapy: A three-level model of outcome variability [J]. *Journal of Consulting and Clinical Psychology*, 2019, 87 (4): 345-356.
- [13] Goldberg S B, Rousmaniere T, Miller S D, et al. Do psychotherapists improve with time and experience? A longitudinal analysis of outcomes in a clinical setting [J]. *Journal of Counseling Psychology*, 2016, 63 (1): 1-11.
- [14] 秦旻. 门诊心理治疗的效果评价研究 [D]. 西南大学, 2010.
- [15] Jacobson N S, Truax P. Clinical significance: A statistical approach to defining meaningful change in psychotherapy research [J]. *Journal of Consulting and Clinical Psychology*, 1991, 59 (1): 12-19.
- [16] 李钰静. 中文版 OQ-45. 2 信效度研究及初步应用 [D]. 浙江理工大学, 2010.
- [17] 李钰静, 骆宏. 心理咨询效果评估量表中文版的信度和效度 [J]. *中国心理卫生杂志*, 2009, 23 (2): 105-107.
- [18] Lambert M J, Hansen N B, Harmon S C. Outcome Questionnaire System (The OQ System): Development and practical applications in healthcare settings. In Barkham M, Hardy GE, Mellor-Clark J. *Developing and Delivering Practice-based Evidence: A Guide for the Psychological Therapies* [M]. Chichester, United Kingdom: Wiley, 2010: 141-154.
- [19] Duan C, Hill C E, Jiang G, et al. Therapist directives: Use and outcomes in China [J]. *Psychotherapy Research*, 2012, 22 (4): 442-457.
- [20] Bohart A C, Wade A G. The client in psychotherapy. In Lambert MJ. *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change, Sixth Edition* [M]. Hoboken, New Jersey: Wiley, 2013: 219-257.
- [21] Clark D M, Fairburn C G, Wessely S. Psychological treatment outcomes in routine NHS services: A commentary on Stiles et al. (2007) [J]. *Psychological Medicine*, 2008, 38 (5): 629-634.
- [22] Stiles W B, Barkham M, Mellor-Clark J, et al. Routine psychological treatment and the Dodo verdict: A rejoinder to Clark et al. (2007) [J]. *Psychological Medicine*, 2008, 38 (6): 905-910.
- [23] 江光荣, 夏勉. 心理求助行为: 研究现状及阶段-决策模型 [J]. *心理科学进展*, 2006, 14 (6): 888-894.
- [24] Lambert M J, Whipple J L, Kleinstäuber M. Collecting and delivering progress feedback: A meta-analysis of routine outcome monitoring [J]. *Psychotherapy*, 2018, 55 (4): 520-537.
- [25] 余壮, 孙启武, 江光荣, 等. 心理治疗中的正式反馈 [J]. *心理科学进展*, 2017, 25 (7): 1197-1207.

The Effectiveness of Counseling Delivered at A University Mental Health Service Center: Practice-based Evidence

Yan Yupeng¹ Jiang Guangrong¹ Ying Mengting² Yu Lixia¹ Lin Xiubin¹
Li Danyang¹

1. School of Psychology, Central China Normal University, Key Laboratory of Human Development and Mental Health of Hubei Province, Key Laboratory of Adolescent Cyberpsychology and Behavior (Ministry of Education), Wuhan;

2. Department of Psychology, College of Health Science, Wuhan Sports University, Wuhan;

3. Mental Health Service Center, Central China Normal University

Abstract: Objectives: To investigate the effectiveness of counseling at a university mental health center, therefore providing practice-based evidence in China. Methods: The data of this study were collected from a research database that was established ten years ago regarding counseling process-outcome in a university mental health center. One hundred and three university students who participated in this study were reduced to a sample of 63 clients (mean age = 21.75, $SD = 2.28$; female 79.4%; average number of session = 7.05, $SD = 3.53$) after removing some data that couldn't be used to calculate pre-post outcomes due to missing data or ending too early (i.e. terminating counseling after just one or two sessions). A clinical sample of 31 clients (average number of session = 7.45, $SD = 3.71$) was generated according to the clinical cut-off score (a score of 63) on the Chinese version of Outcome Questionnaire-45.2 (OQ-45.2). The effect size and the clinically significant change status of the samples were benchmarked against that of two clinical samples from two American university counseling centers. Results: (1) This university mental health center obtained a medium pre-post effect size ($d = 0.57$) based on the total sample and a large pre-post effect size ($d = 1.22$) based on the clinical sample, which was larger than that of the available clinical sample from an American university counseling center; (2) According to the clinically significant change status, 14 clients (22.2%) recovered, 9 clients (14.3%) improved, 36 clients (57.1%) didn't change and 4 clients (6.3%) deteriorated, which was similar to the available general sample from another American university counseling center with a non-significant difference. However, the rates of recovery of the two samples differed significantly ($p < 0.05$) when these two samples were divided into the recovered and the unrecovered group. Conclusion: Counseling delivered at this university mental health center seemed to be effective which was not worse than two American samples and even better in some aspects, which should be considered cautiously due to the data limitation.

Key words: University mental health center; Counseling effectiveness; Practice-based evidence; Benchmarking