**抑郁问卷的异质性：基于对27个抑郁测量问卷的内容分析**

汪浩远1, 胡孟真1, 田柳青1, 刘伟彪1, 安媛媛1, 李瑛2, 胡传鹏1#

(1南京师范大学心理学院; 2首都医科大学附属北京儿童医院精神科)

**Assessing the heterogeneity of 27 scales for measuring depression**

Haoyuan Wang, Mengzhen Hu, XXX

**CrediT Author Statement**

**Wang Hao Yuan：**Conceptualization, Data curation, Investigation, Writing - Original Draft. **Hu Meng Zhen：**Data curation, Visualization, Investigation. **Tian Liu Qing:** Investigation. **Liu Wei Biao:** Investigation. **Yuanyuan An:** Writing- Reviewing and Editing. **Ying Li:** Writing- Reviewing and Editing. **Hu Chuan-Peng：**Supervision**,** Project administration, Writing- Reviewing and Editing.

Corresponding author: Hu Chuan-Peng, email: hcp4715@hotmail.com

# 1. Introduction

[抑郁障碍的严重性（患病率、社会成本等），青少年期和成年早期是关键的时间。]

Depression can induce considerable distress, hinder social functioning and economic productivity, and result in premature mortality. Its effects extend significantly to individuals close to the affected person and have broader societal implications(Herrman et al., 2022). In China, major depression ranked as the second leading cause of disability(Yang et al., 2013). According to the "China National Mental Health Development Report (2021-2022)" released by Institute of Psychology of the Chinese Academy of Sciences, Chinese people have a depression detection rate of 10.6%(傅小兰 & 张侃, 2023). Depression can also bring about economic burdens. The economic burden of major depressive disorder among US adults increased from $236.6 billion to $326.2 billion (Greenberg et al., 2021).The adolescent and early adulthood periods are crucial times, with the depression risk detection rate reaching as high as 24.1% in the age group of 18-24, significantly higher than other age groups(傅小兰 & 张侃, 2023).

[上述关于抑郁障碍的患病率大部分来自于自评量表，在抑郁症的研究中广泛使用，也是各类关于青少年与学生政策的基础（中国、世界范围内大规范调查的的数据）。]

The reported rates of depression in China, as previously mentioned, are derived from self-report instruments, specifically utilizing the CESD-9. Furthermore, numerous extensive studies conducted in China persist in utilizing self-report measures to assess and determine depression detection rates. Specific details can be found in Table 1.Moreover, the results of these large-scale surveys form the policy backdrop for the prevention and treatment of depression in China(Ma et al., 2023).The Chinese National Health Commission has initiated an exploratory program for the prevention and treatment of depression, proposing the use of the PHQ-9 scale for conducting depression screening.

Table1：中国大规模调查

|  |  |  |
| --- | --- | --- |
| 文章名 | 调查范围 | 测量方式 |
| 中国国民健康心理发展报告 | 全国31个省，191347人 | CESD-9 |
| 2021和2022年中国居民抑郁和焦虑症状及其共患的相关因素 | 全国23个省，41426人 | PHQ-9 |
| Prevalence of depressive disorders and treatment in China:  a cross-sectional epidemiological study | 全国31个省，28140人 | QIDS-SR |

[但作为解决抑郁问题第一步的测量与诊断问题，目前较为忽视，列举Fried的工作。不同抑郁问题可能在测量不同的内容。Fried（2017年）的研究表明，不同的抑郁量表不能互相替代使用，这给抑郁研究带来了挑战。]

However, the issue of measurement and diagnosis, which serves as the first step in addressing depression, is currently relatively overlooked. Fried (2017) pointed out that there is an untested assumption, that various depression assessment tools can be interchangeable measures of the severity of depression. He analyzed seven common depression scales and identified a total of 52 different depressive symptoms. He found that the overall overlap of item content in the questionnaires was very low, indicating that the conventional practice of using scales as interchangeable measures of depression severity is problematic. This approach may pose a significant threat to the generality and replicability of depression research.

[在中国，同样存在着抑郁自评量表数量繁多，被广泛使用的量表也不在少数，有潜在的重大影响（如影响到心理健康问题的检出率）。]

In China, research on depression primarily relies on the measurement of scales, with many widely used depression scales. This practice may have potential significant implications. For instance, recent meta-analysis results on the detection rates of mental health issues among Chinese students indicate a significant moderating effect of measurement tools, particularly in the context of depression(Huang et al., 2022; YU Xiaoqi et al., 2022; Zhang et al., 2022).Furthermore, even with the same scale, there are substantial variations in the detection rates obtained for the SCL-90 under different screening criteria(Chen et al., 2022).

[本研究准备怎么做]

Within the realm of psychology, self-report scales play a pivotal role in quantifying abstract constructs, such as depression and anxiety. Conclusions drawn in the context of depression, for instance, hinge significantly on the scores derived from a specific scale. Consequently, evaluating the degree of overlap among scales measuring the same variable becomes imperative to gauge the reliability of our knowledge base. Considering the extensive landscape of depression research in China and its societal impact, it becomes essential to scrutinize the actual overlap between depression scales employed in Chinese studies. This study aims to scrutinize the depression scales utilized in Chinese research with the objectives of (1) assessing the content overlap among these scales, (2) examining the extent to which DSM-5 symptoms are measured across diverse scales, and (3) offering a precise description of the content appraised by different depression scales.

# 2. Methods

We took three steps to extract symptoms from all scales that measure depressions among the Chinese students population (see Fig 1 for the flowchart). Firstly, we identified all scales that has been used for screening depression. Secondly, we identified unique symptoms of each scale. Thirdly, we compared the symptoms across all scales. The latter two steps followed Fried (2017)but with modifications (see details below).

## 2.1 Identify and screen scales

We identified scales that measure depression from four recent meta-analyses which synthesized the prevalence of different mental health problems among four Chinese students populations: elementary school, middle school, high school, and college (于晓琪等, 2022; 黄潇潇等, 2022; 张亚利等, 2022; 陈雨濛等, 2022). We extracted all papers included in these meta-analyses and selected scales that were used for screening depression. In total, we identified 34 scales from 470 articles from all articles included in these four meta-analysis.

We then screened versions of scales and identified the most valid version for later analysis (See supplementary material for details). If a scale has multiple versions, we choose one of these version based on the following criteria: (1) If the reference(s) of the scale cited in articles was in English, we search for the Chinese versions because all participants in the above mentioned articles are Chinese students; (2) If the reference(s) of the scale included both English and Chinese versions of the scale, we included the Chinese version; (3) If there were multiple Chinese versions and the latter one(s) improved the previous ones for methological reasons, we choosed the latter version; (4) If there were multiple Chinese versions and no obvious methodological advances were reported, we included the one explicitly stated symptom names; (5) If all other condition were equal, we selected the most frequently cited one. For instance, CES-D was first translated by汪向东等(1999), which was used by 39 of all 470 papers included in the four menta-analyses. However, 章婕等(2010)improved the translation of the 20th item, thus, we chose the revised version by 章婕等(2010).

## 2.2 Identify symptoms within scales

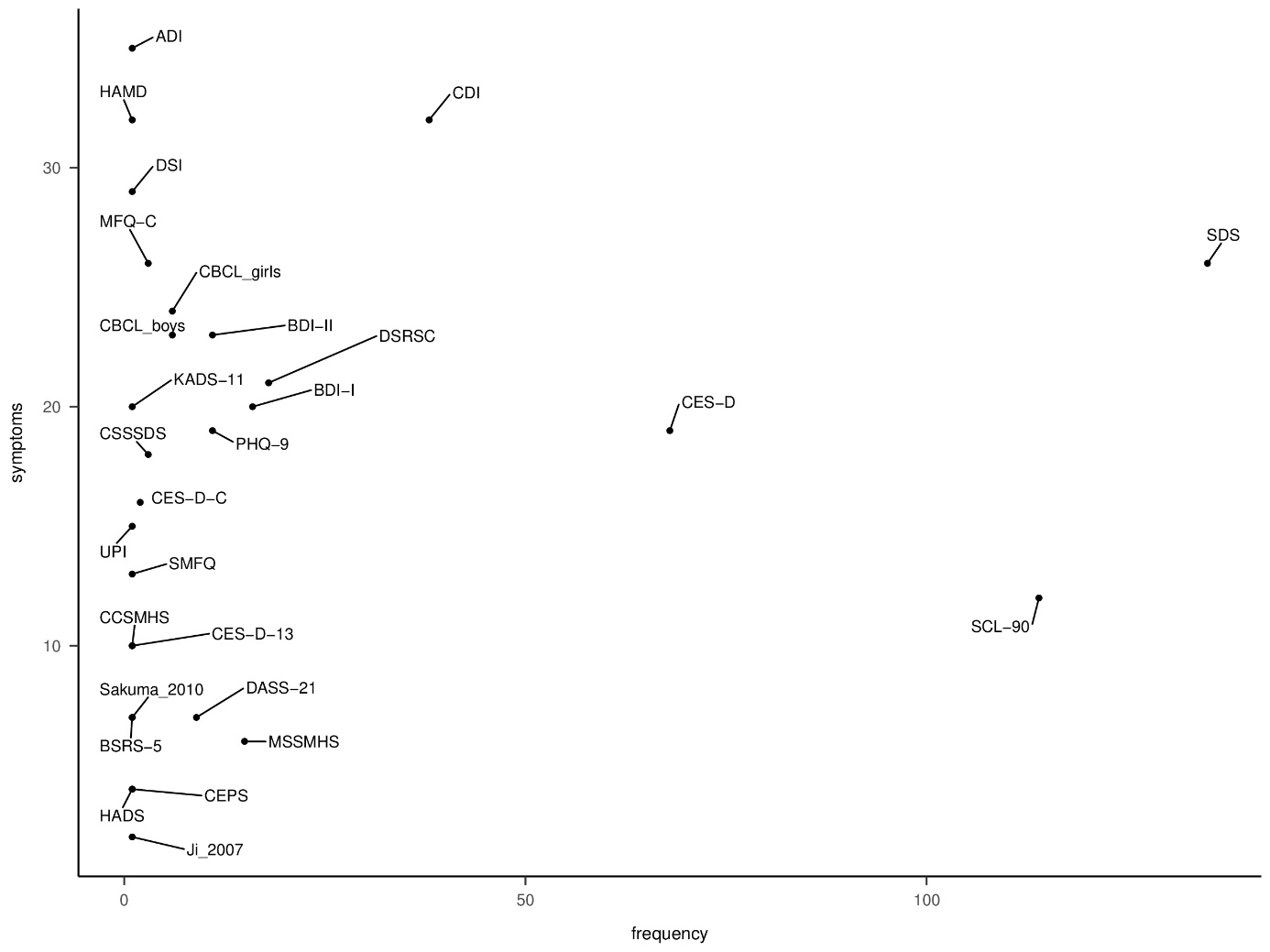
In this phase, four trained coders focused on identifying items that assessed identical or similar symptoms within each scale. The process proceeded as follows: First, the four coders independently identified and consolidated items within each scale. Then, they formed two pairs, with each pair reviewing their results and resolving any discrepancies. Subsequently, the two pairs of coders cross-checked their findings and collaboratively addressed any remaining inconsistencies, seeking guidance from the corresponding author when necessary. The final, consolidated version of the results underwent independent verification by a clinically certified doctor (co-author \*\*\*), who made necessary revisions to ensure accuracy.

## 2.3 Compare symptoms across different scales

In this step, the same four coders compared symptoms across all scales. The procedure was the same as the identification of symptoms within scales: indepenedent individual coding, discussion by pairs, cross-checking between pairs, discussion with the corresponding author, and verification by a clinically certificated doctor.

The coding processes were as follow. We first pooled symptoms from all scales together and identified unique symptoms. In this step, we retained both compound symptoms and specific symptoms, as in Fried (2017). Compound symptoms are symptoms that include a range of related symptoms, whereas specific symptoms are more concrete and describe specific patterns. For example, “appetite changes” is a compound symptom, it includes two specific symptoms: “appetite increased” and “appetite decreased”. And all three symptoms were treated as unique symptoms. We employed an approach that maximize the amount of different symptoms. More specifically, if the items describe similar symptom using different words and that the words have significantly different meanings under the Chinese context, we treat them as belonging to the same compound symptom but are different specific symptoms. For instance, there are many different words for describe depressed mood in different scales, we used 'depressed moods' as the compound symptom but distinguish different specific symptoms such as: 'blue', 'low mood', 'sad', and 'anhedonia'. This approach is slightly different from Fried (2017), where he coded all these item as a specific symptom 'Sad moods'.

Then, we coded each scale on all unique symptoms. More specifically, a scale was coded as “0” on a symtom if it does not have items that measures this symptom. If a scale has an item that directly measures a symptom, compound or specific, it was coded as 2 on that symptom. Note if a scale has an item measures a compound symptom, then, this scale not only had a score of 2 on that compound symptom but also has a score of 1 on each of the specific symptoms under this compound symptom. For instance, the CDI has no item for 'Depressed mood' , we assigned “0” for CDI on this symptom. For compound symptom “appetite change”, CDI has an item directly measures this symptom and was coded as 2 on this symptom. Importantly, even CDI does not have item for specific symptoms of “appetite changes”, i.e., 'appetite increased' and 'appetite decreased', it was coded as 1 on both two specific symptoms (see Supplementary Materials and Figure 2 for details). However, if the item measures a specific symptom under a compound symptom, this scale was coded “2” on that specific symptom but also coded as “0” on the corresponding compound symptom.

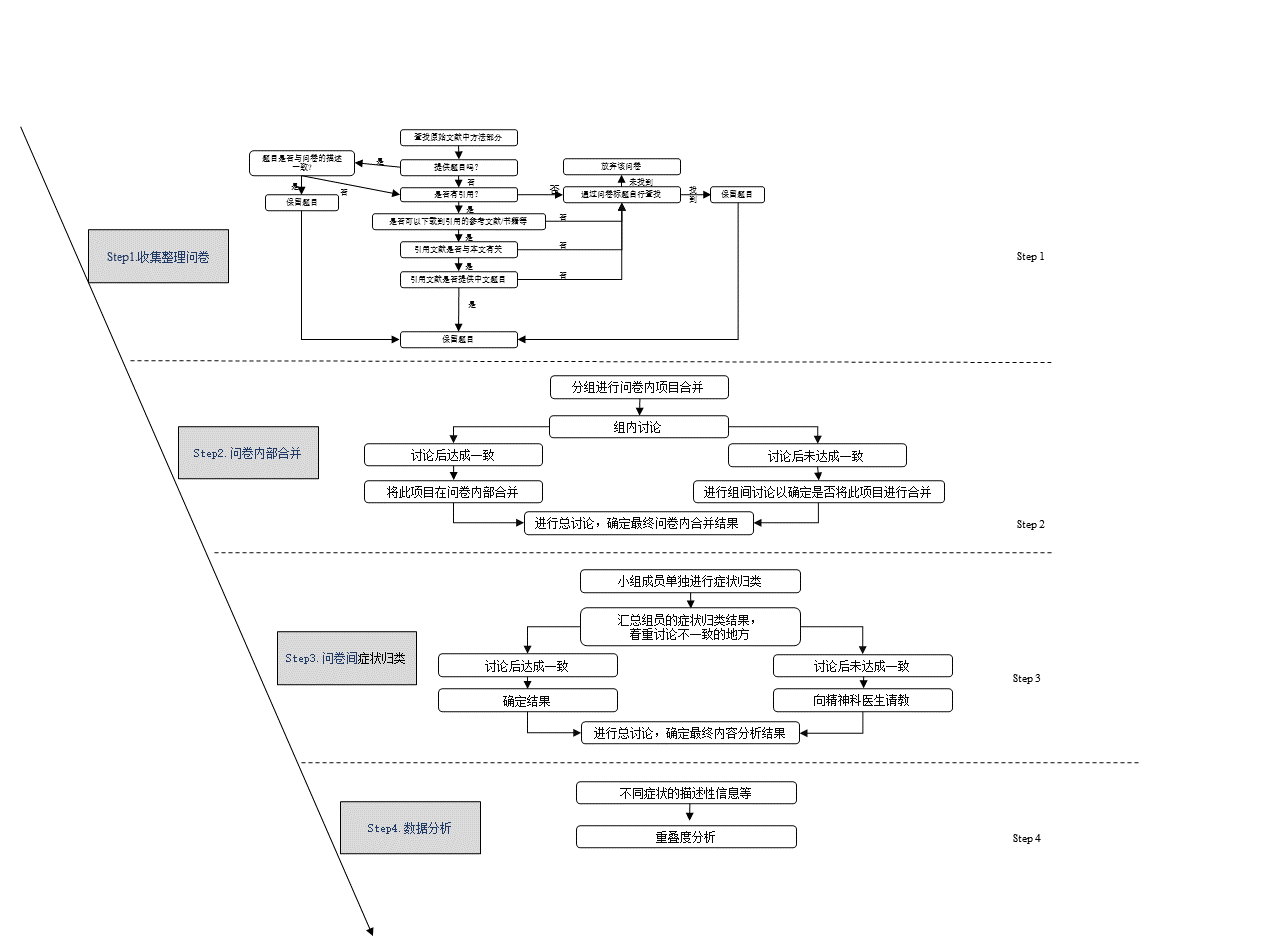


**Figure 2***. Number of articles that used each scale for screening depression.* Total number of articles were based on those included in four meta-analyses (XXX). Scales with \* were excluded from our analyses.

## 2.4 Data analyses

We reported the descriptive summary of scales as well as the symptoms within scale. We also highlightd symptoms that are used in DSM-5 for diagonosis of depression. More specifically, there are 28 symptoms are DSM-5 symptom, derived from the nine symptoms in DSM-5 and all their specific symptoms, see Fried & Nesse, (2015) for details.

We analyzed shared and distinct symptom across scales. We used Jaccard index for the degree of overlap between different scales (Fried, 2017). The formal of Jaccard index is s/(u1 + u2 + s), where "s" represents the number of items shared by two scales, and "u1" and "u2" denote the number of items that are exclusively present in each of the two scales. Jaccard index ranges from 0 (no overlap among scales) to 1 (complete overlap). We interpreted Jaccard index as in Fried (2017) guidelines: very weak 0.00–0.19, weak 0.20–0.39, moderate 0.40–0.59, strong 0.60–0.79, and very strong 0.80–1.0. We explored the relationship between the mean Jaccard coefficient and the scale length across 27 scales by employing Spearman correlation.



**Figure 1**

*Research flowchart*

# 3 Result

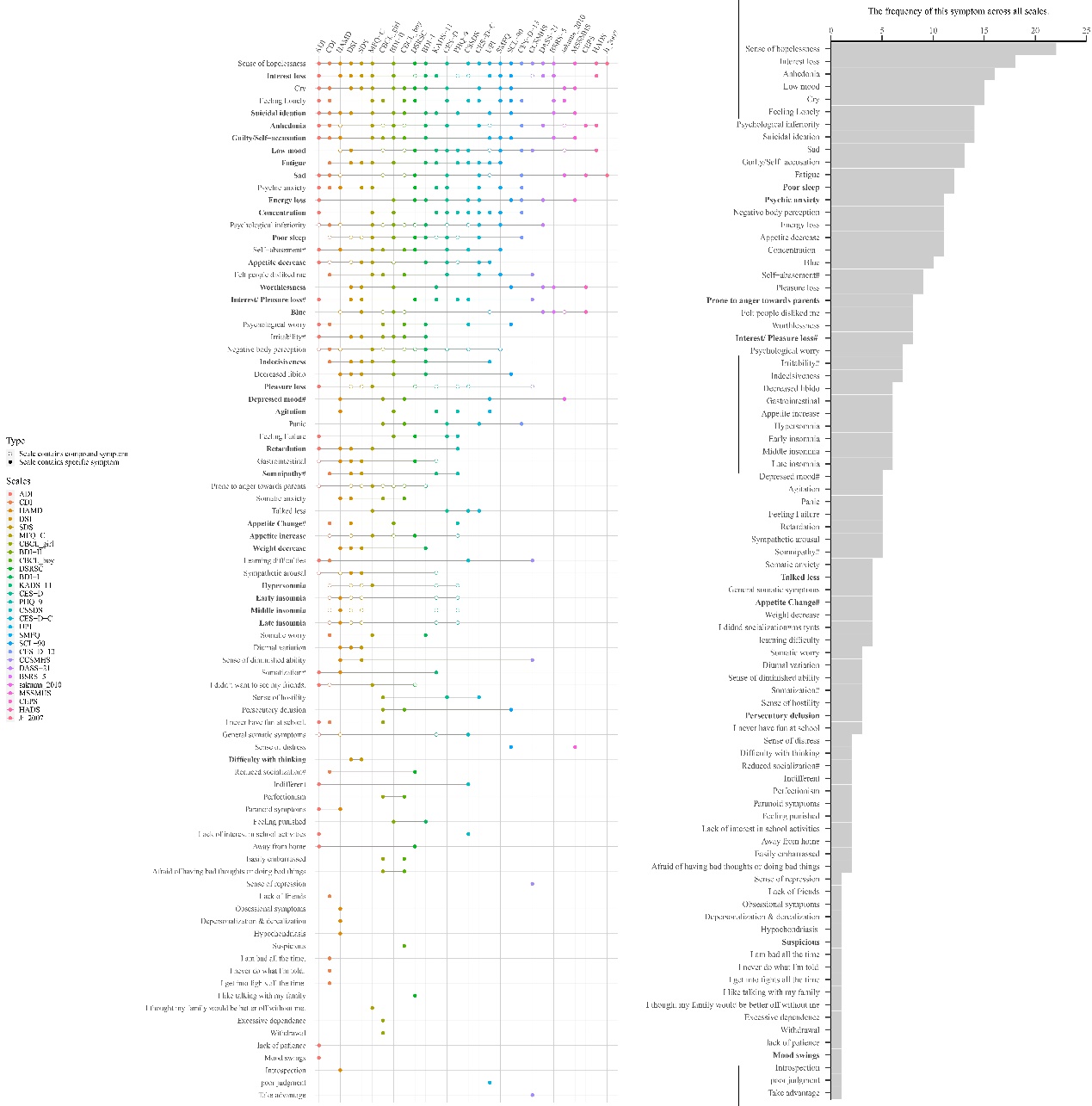
# 3.1 A summary of scales

Among all papers included in these four meta-analyses (citations), 470 reported depression. We identified 34 unique scales in these articles. Among all these scales, the items of four scales, the Mini International Neuropsychiatric Interview for children and adolescents (Mini-KID), WHO-CIDI 3.0, Psychological Health Inventory (PHI), and the Symptom Checklist 45, were not findable. The items of the other two scales, the Beck Depression Inventory (Zhang Yuxin Revised Edition) and Short Depression Scale, were not available either because of unidentifiable citations. These six scales were excluded from furthere analyses. The items used in 'Gu & Chen (2020) 'and 'Ji (2007)' were identifical but in different languages, thus we regarded these two studies used the same scale referred them as 'Ji (2007)'. Also, the boys’ and girls’ version of the Child Behavior Checklist (CBCL) were treated as one scale, CBCL. In short, 27 scales were included in the current study. See figure 2 for the number of usages of these scales among all 470 empirical papers in the meta-analyes (CITATIONS). The seven most frequently used scales in this study are SDS, SCL-90, CES-D, CDI, DSRSC, BDI-I, and MSSMHS, among which CES-D and SDS are consistent with the scales selected by Fried (2017).

# 3.2 Items and symptoms within scales

For the 27 scales included, there are 425 items in total. We merged 73 items into 31 symptoms when coding items within scales. Also, there were two cases where one item measured two symptoms. The item from Ji (2007), "During the past 12 months, did you ever feel so sad or hopeless almost every day for two weeks or more in a row that you stopped doing your usual activities?" meausres both 'sad' and 'Sense of hopelessness'. And the 8th item from PHQ-9, "Actions or speech slowed down to a noticeable extent, or conversely—feeling restless or agitated, being unable to sit still, more than usual", measures both 'Agitation' and 'Retardation'. Thus, a total of 385 symptoms were identified across different scales (See supplementary materials for number of items and symptoms of each included scale).

The comparison of 385 symptoms across 27 scales resulted in unique 84 symptoms (see Figure 3). Among these, eight are compound symptoms, including *Depressive mood, Irritability*, *Self-abasement*, *Interest/pleasure loss*, *Somatization*, *Appetite changes*, *Somnipathy*, and *Reduced socialization* (See supplementary material for details). Among all 27 scales, 19 of them did not include any idiosyncratic symptoms. For the other eight scales the rate of idiosyncratic symptom varied from 3.9% to 22.2%. Interestingly, all scales include symptoms that were not used for diagnosis of depression in DSM-5. Among them, DSI has the highest proportion of DSM-5 symptoms for depression, 71.42% of the total nine DSM-5 depression symptoms. Please see the supplementary matierals for detailed information.



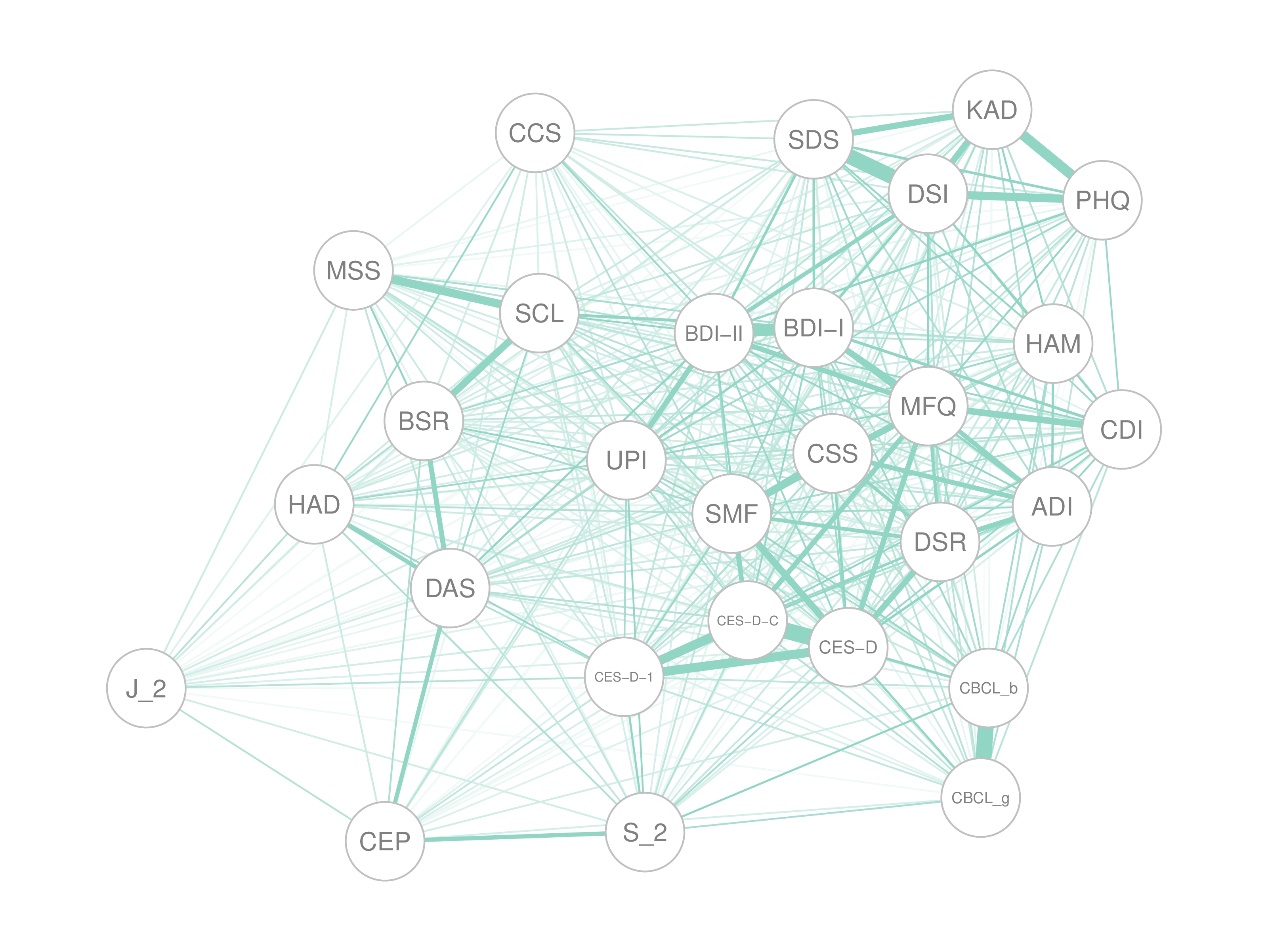
**Figure 3***. Content Overlap Across* *Twenty-seven Depression Scales.* Each row represents a symptom, each column represents a scale.If a scale measures a symptom, then there is a dot or a circle on that row. The former represent compound symptoms and the latter represent specific symptoms. Symptoms in bold font are from DSM-5. ADI: Adolescent Depression Inventory, CDI: Children's Depression Inventory, HAMD: Hamilton Depression Rating Scale for Depression, DSI: Depression Status Inventory, SDS: Self- Rating Depression Scale, MFQ-C: Mood and Feelings Questionnaire, CBCL: Child Behavior Checklist, BDI-II：Beck Depression Inventory-II, DSRSC: Depression Self-rating Scale for Children, BDI-I: Beck Depression Inventory, KADS-11: Kutcher Adolescent Depression Scale, CES-D: The Center for Epidemiological Studies Depression Scale, PHQ-9: Patient Health Questionnaire-9 items, CSSDS: China Middle school students' depression scale, CES-D-C: Center for Epidemiologic Studies Depression Scale for Children, UPI: University Personality Inventory, SMFQ: Short Mood and Feelings Questionnaire, SCL-90: Symptom Checklist 90, CES-D-13: Short version of Center for Epidemiologic Studies Depression Scale, CCSMHS: Chinese College Student Mental Health Scale, DASS-21: The Depression Anxiety Stress Scale, BSRS-5: Brief Symptom Rating Scale, Sakuma\_2010: Sakuma et al.(2010) self-designed questionnaire, MSSMHS: Middle-school students Mental Health Scale, CEPS: China Education Panel Survey, HADS: Hospital Anxiety and Depression Scale, Ji\_2007: Comprehensive Survey Report on Health-Related/Risk Behaviors among Chinese Adolescents.The right side of the figure is *Number of symptoms that appear across combinations of scales.* There are 18 symptoms that appear only in one scale; these symptoms are referred to as idiosyncratic symptoms.

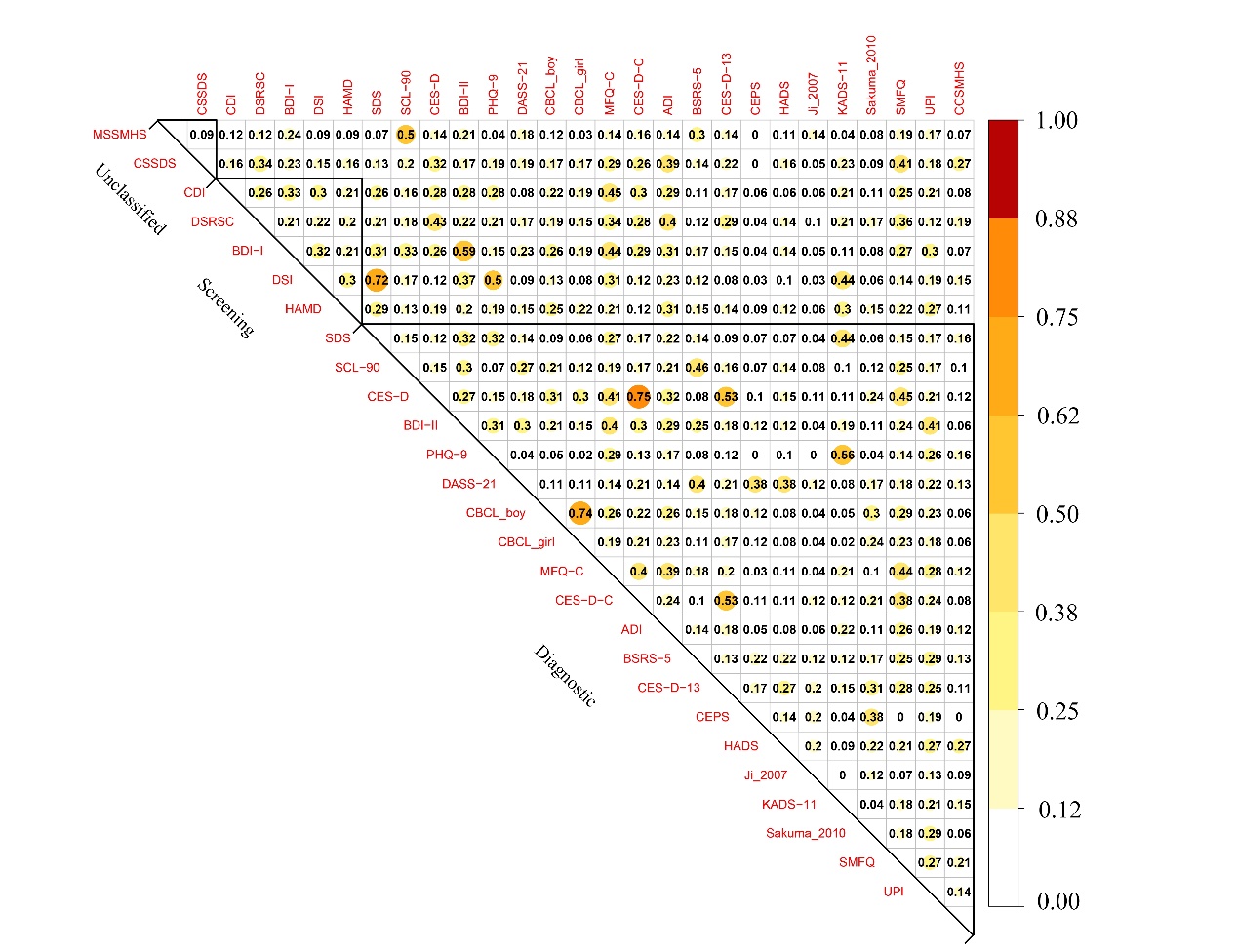
# 3.3 Symptoms across scales

Among the 84 symptoms, 18 (21.43%) were idiosyncratic symptoms and only appeared in one scale. None of the symptoms were present on all scales. The most frequently symptom, appeared in 22 out of 27 scales, was *Sense of hopelessness*. The second most measured symptom, 18 out of 27, was *Interest loss*. Note that *markedly diminished interest or pleasure*, a key symptom of diagnosis of major depression in DSM-5, is splitted into two specific symptoms: *interest loss* and *pleasure loss* in this study*.* We found *Pleasure loss* was observed less frequently than *Interest loss*, being measured in 9 out of 27 scales. Another frequently measured symptom is the compound symptom *Depressed mood*, which was directly measured in 5 scales. However, this compound symptom include several specific symptoms: *blue* appeared in 10 scales, *low mood* in 15 scales, *sad* in 13 scales, and *anhedonia* in 16 scales*.* Combined as a cluster, depressed mood and related specific symptom are the most frequently measured.

The degree of overlap between scales was calculated using the Jaccard coefficient. The mean overlap across all scales was 0.19, range from 0.09 to 0.25, indicating a very low level of similarity between these scales (see Figure 5). CES-D has the highest average degree of overlap with other scale. The highest overlap, 0.75, appeared between two versions of CES-D: CES-D for adults and CES-D-C for chidlren. The second highest overlap, 0.72, was between DSI and SDS. Many scales that have no overlap with each other. For example, there was no overlap between China Education Panel Survey with Middle-school students Mental Health Scale, PHQ-9, CSSDS, SMFQ, CSSMHS. Note that because Ji\_2005 has only one item that measures two symptoms, it has no overlap with PHQ-9, KADS-11.

We found a correlation between the mean Jaccard coefficient of each scale and the scale length, *r* = 0.54, 95% CI [0.14, 0.78]. Similarly, the correlation between mean overlap of scales and the number of captured symptoms is *r* = 0.70, 95% CI [0.39, 0.87]. These findings suggests that longer scales exhibit increased overlap with other scales, thus demonstrating enhanced representativeness.





**Figure 5.** *Overlap of item content of 27 depression scales.* ADI: Adolescent Depression Inventory, CDI: Children's Depression Inventory, HAMD: Hamilton Depression Rating Scale for Depression, DSI: Depression Status Inventory, SDS: Self- Rating Depression Scale, MFQ-C: Mood and Feelings Questionnaire, CBCL: Child Behavior Checklist, BDI-II：Beck Depression Inventory-II, DSRSC: Depression Self-rating Scale for Children, BDI-I: Beck Depression Inventory, KADS-11: Kutcher Adolescent Depression Scale, CES-D: The Center for Epidemiological Studies Depression Scale, PHQ-9: Patient Health Questionnaire-9 items, CSSDS: China Middle school students' depression scale, CES-D-C: Center for Epidemiologic Studies Depression Scale for Children, UPI: University Personality Inventory, SMFQ: Short Mood and Feelings Questionnaire, SCL-90: Symptom Checklist 90, CES-D-13: Short version of Center for Epidemiologic Studies Depression Scale, CCSMHS: Chinese College Student Mental Health Scale, DASS-21: The Depression Anxiety Stress Scale, BSRS-5: Brief Symptom Rating Scale, Sakuma\_2010: Sakuma et al.(2010) self-designed questionnaire, MSSMHS: Middle-school students Mental Health Scale, CEPS: China Education Panel Survey, HADS: Hospital Anxiety and Depression Scale, Ji\_2007: Comprehensive Survey Report on Health-Related/Risk Behaviors among Chinese Adolescents. Mean overlap is detailed in the supplementary materials.

# 4 讨论

[第一段: 总结结果]

研究分析了多个常用的抑郁量表，并发现它们在捕捉抑郁症状方面存在较大的异质性。这些量表之间的项目内容重叠度较低。这对于使用和解释相关数据有重要的启示作用。

[第二段: 与已有研究进行比较]

Fried (2017).的研究中Ces-d与其他量表有着最差的平均重叠率（0.27），而Ces-d在本研究中表现出最高的平均重叠率（0.25），可能是由于：1、本研究中除Ces-d以外还纳入了Ces-d简版以及Ces-d儿童版。2、在他的研究中Ces-d有33%的独特症状，而本研究中，随着纳入的量表增加，Ces-d的独特症状为0。

[第三段: 重新阐述问题的严重性，尤其是与大规模调查进行结合]

抑郁量表的异质性来源可能是由于构念的不清晰，因此，在合并来自不同量表的研究结果时（例如元分析），特别是那些几乎没有重叠的量表时，可能不明智。目前仍普遍认为量表可以互相替代，研究结果表明量表之间异质性非常强。因此在推广结果中应该强调这是某个量表的结果。

[第四段：如何解读结果，call for better standards]

不能认为低重叠就是糟糕的量表，高重叠就是好的量表。选择测量工具时应考虑多种因素，并根据评估的目的和目标进行选择。

[第五段: 展望未来]

**参 考 文 献**

Chen Y., Zhang Y., Yu G., & 俞国良. (2022). Prevalence of mental health problems among college students in mainland China from 2010 to 2020: A meta-analysis. *Advances in Psychological Science*, *30*(5), 991–1004. https://doi.org/10.3724/SP.J.1042.2022.00991

Fried, E. I. (2017). The 52 symptoms of major depression: Lack of content overlap among seven common depression scales. *Journal of Affective Disorders*, *208*, 191–197. https://doi.org/10.1016/j.jad.2016.10.019

Fried, E. I., & Nesse, R. M. (2015). Depression sum-scores don’t add up: Why analyzing specific depression symptoms is essential. *BMC Medicine*, *13*(1), 72. https://doi.org/10.1186/s12916-015-0325-4

Greenberg, P. E., Fournier, A.-A., Sisitsky, T., Simes, M., Berman, R., Koenigsberg, S. H., & Kessler, R. C. (2021). The Economic Burden of Adults with Major Depressive Disorder in the United States (2010 and 2018). *Pharmacoeconomics*, *39*(6), 653–665. https://doi.org/10.1007/s40273-021-01019-4

Herrman, H., Patel, V., Kieling, C., Berk, M., Buchweitz, C., Cuijpers, P., Furukawa, T. A., Kessler, R. C., Kohrt, B. A., Maj, M., McGorry, P., Reynolds, C. F., Weissman, M. M., Chibanda, D., Dowrick, C., Howard, L. M., Hoven, C. W., Knapp, M., Mayberg, H. S., … Wolpert, M. (2022). Time for united action on depression: A Lancet–World Psychiatric Association Commission. *The Lancet*, *399*(10328), 957–1022. https://doi.org/10.1016/S0140-6736(21)02141-3

Huang X., Zhang Y., & Yu G. (2022). Prevalence of mental health problems among primary school students in Chinese mainland from 2010 to 2010:A meta-analysis. *Advances in Psychological Science*, *30*(5), 953–964. https://doi.org/10.3724/SP.J.1042.2022.00953

Ma, J., Zhou, H., Fu, Q., & Lu, G. (2023). Facilitators and barriers in the development and implementation of depression prevention and treatment policies in China: A qualitative study. *BMC Public Health*, *23*(1), 276. https://doi.org/10.1186/s12889-023-15201-0

Yang, G., Wang, Y., Zeng, Y., Gao, G. F., Liang, X., Zhou, M., Wan, X., Yu, S., Jiang, Y., Naghavi, M., Vos, T., Wang, H., Lopez, A. D., & Murray, C. J. (2013). Rapid health transition in China, 1990–2010: Findings from the Global Burden of Disease Study 2010. *The Lancet*, *381*(9882), 1987–2015. https://doi.org/10.1016/S0140-6736(13)61097-1

YU Xiaoqi, ZHANG Yali, & YU Guoliang. (2022). Prevalence of mental health problems among senior high school students in mainland of China from 2010 to 2020: A meta-analysis. *Advances in Psychological Science*, *30*(5), 978. https://doi.org/10.3724/SP.J.1042.2022.00978

Zhang Y., Jin J., & Yu G. (2022). Prevalence of mental health problems among junior high school students in Chinese mainland from 2010 to 2020: A meta-analysis. *Advances in Psychological Science*, *30*(5), 965–977. https://doi.org/10.3724/SP.J.1042.2022.00965

傅小兰 & 张侃. (2023). *心理健康蓝皮书 中国国民心理健康发展报告(2021-2022)* (1st ed.). 社会科学文献出版社.

汪向东, 王希林, & 马弘. (1999). *心理卫生评定量表手册*. 中国心理卫生杂志社.

章婕, 吴振云, 方格, 李娟, 韩布新, & 陈祉妍. (2010). 流调中心抑郁量表全国城市常模的建立. 中国心理卫生杂志, *24*(2), 139–143.