



Article

Psychological Well-Being of Left-Behind Children in China: Text Mining of the Social Media Website Zhihu

Yuwen Lyu ^{1,2}, Julian Chun-Chung Chow ³ , Ji-Jen Hwang ^{4,5}, Zhi Li ⁶, Cheng Ren ³ and Jungui Xie ^{7,*}

¹ School of Economics and Statistics, Guangzhou University, Guangzhou 511400, China; 1111764007@e.gzhu.edu.cn

² Center for Information Technology Research in the Interest of Society and the Banatao Institute, University of California, Berkeley, CA 94720, USA

³ School of Social Welfare, University of California, Berkeley, CA 94720, USA; jchow99@berkeley.edu (J.C.-C.C.); cheng.ren@berkeley.edu (C.R.)

⁴ Center for General Education, Chung Yuan Christian University, Taoyuan 320314, Taiwan; jhwang29@gmu.edu

⁵ School of Policy and Government, George Mason University, Arlington, TX 22030, USA

⁶ School of Information, University of California, Berkeley, CA 94720, USA; zhili@berkeley.edu

⁷ School of Public Administration, Guangzhou University, Guangzhou 511400, China

* Correspondence: junguixie99@gmail.com

Abstract: China's migrant population has significantly contributed to its economic growth; however, the impact on the well-being of left-behind children (LBC) has become a serious public health problem. Text mining is an effective tool for identifying people's mental state, and is therefore beneficial in exploring the psychological mindset of LBC. Traditional data collection methods, which use questionnaires and standardized scales, are limited by their sample sizes. In this study, we created a computational application to quantitatively collect personal narrative texts posted by LBC on Zhihu, which is a Chinese question-and-answer online community website; 1475 personal narrative texts posted by LBC were gathered. We used four types of words, i.e., first-person singular pronouns, negative words, past tense verbs, and death-related words, all of which have been associated with depression and suicidal ideations in the Chinese Linguistic Inquiry Word Count (CLIWC) dictionary. We conducted vocabulary statistics on the personal narrative texts of LBC, and bilateral *t*-tests, with a control group, to analyze the psychological well-being of LBC. The results showed that the proportion of words related to depression and suicidal ideations in the texts of LBC was significantly higher than in the control group. The differences, with respect to the four word types (i.e., first-person singular pronouns, negative words, past tense verbs, and death-related words), were 5.37, 2.99, 2.65, and 2.00 times, respectively, suggesting that LBC are at a higher risk of depression and suicide than their counterparts. By sorting the texts of LBC, this research also found that child neglect is a main contributing factor to psychological difficulties of LBC. Furthermore, mental health problems and the risk of suicide in vulnerable groups, such as LBC, is a global public health issue, as well as an important research topic in the era of digital public health. Through a linguistic analysis, the results of this study confirmed that the experiences of left-behind children negatively impact their mental health. The present findings suggest that it is vital for the public and nonprofit sectors to establish online suicide prevention and intervention systems to improve the well-being of LBC through digital technology.

Keywords: left-behind children; personal narrative text; text mining; linguistic analysis; textual analysis; psychological well-being



Citation: Lyu, Y.; Chow, J.C.-C.; Hwang, J.-J.; Li, Z.; Ren, C.; Xie, J. Psychological Well-Being of Left-Behind Children in China: Text Mining of the Social Media Website Zhihu. *Int. J. Environ. Res. Public Health* **2022**, *19*, 2127. <https://doi.org/10.3390/ijerph19042127>

Academic Editor: Paul B. Tchounwou

Received: 27 December 2021

Accepted: 5 February 2022

Published: 14 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

During China's planned economy period, the movement of citizens was controlled, and citizens were allowed to reside only in their registered permanent residence, according

to the household registration (Hukou) system. Since the economic reform of 1978, China has worked towards a market economy and removed barriers deterring mobility, which resulted in the migration of tens of thousands of residents from rural areas to urban ones [1]. According to the latest report from the government, the migrant population in China has reached 290 million, an increase of 11.89% as compared with the 2010 Chinese Census [2]. On the one hand, migrant workers have tremendously contributed to China's economy; on the other hand, the children of migrant workers are facing vital problems that need urgent attention [3]. Left-behind children (LBC) are children younger than 17 years old who stayed in the family's registered permanent residence while their parent or parents migrated to another area for work; LBC need attention and care from the government and society [4]. The 2015 China Family Development Report indicated that LBC in rural areas accounted for 35.1% of the total population of children [5]. According to the National Information Management System developed by the Ministry of Civil Affairs, regarding LBC and children in difficulty in rural areas, there were 6.97 million LBC in 2018, a 22.7% decrease from the 9.02 million LBC recorded in 2016 [6]. However, researchers have taken the Sixth National Population Census (2010) of the People's Republic of China as sample data, and have estimated the number of LBC to be 61.02 million [7]. Irrespective of the method used, i.e., a recorded number or an estimated number, the population of LBC is significant, and their education and psychological development is an urgent social problem for the Chinese government.

In a recent study, researchers reported that male LBC were at higher risk of skipping breakfast, physical inactivity, internet addiction, having ever smoked tobacco, and suicidal ideation, whereas female LBC preferred sugary drinks and had higher risks of alcohol addiction and suicidal ideations than their non-LBC counterparts [8]. In addition, Z. Jia and Tian found that LBC were 2.5 times more likely to suffer from loneliness, and 6.4 times more likely to be very lonely [9]. In another study, conducted using a cross-sectional design with 3254 participants (8 to 17 years old), the results showed that the prevalence of depression was 10.9% among non-LBC and 14.3% among LBC [10]. From the data released in the 2011 Survey on Social Integration of Migrant Children in Wuhan, the results showed that LBC and migrant children had worse psychological and behavioral outcomes than local children, which disappeared with family and school support; this implied that family intervention and education policy reform could improve the psychological and behavioral outcomes of migrant children and LBC through family intervention [11].

However, existing research on LBC mostly consists of data collected and analyzed through questionnaires, mental health standardized scales, and traditional interview methods. Few researchers have focused on big data analytics of social media data. In the current digital age, social network services have become a part of everyone's daily lives; texts, pictures, speech, and videos on social media all reflect the characteristics of people's daily behaviors. As such, big data collected from social media can be very valuable in psychological studies, and therefore offers an important complement to traditional research methods [12]. In addition, text mining can be applied to assess large amounts of textual data that are continuously produced in social networks, by analyzing and discovering high-quality information from texts, to help analysts in decision making [13]. Recent research has applied text mining to the study of children's social welfare. For example, Lyu et al. applied text mining techniques to analyze textual data collected from a social media platform, i.e., the Weibo website [14]. They investigated people's attitudes toward child abuse incidents, as well as the reasons behind the abuse, and provided recommendations to the government to improve child protection policies. S. Li et al. conducted a study on 355 migrant youths in a middle school in Shenzhen [15]. They used a text mining technique, namely natural language analysis, on the children's diary entries to examine the relationship between their language patterns and psychological resilience. Machine learning features were also applied to estimate their level of psychological resilience.

Since 1979, researchers have come to realize that the vocabulary used in oral communication can reflect personal feelings and thoughts. Therefore, a linguistic analysis can be used

as an effective tool to understand a person's psychological state [16]. In 1980, researchers discovered that patients who wrote about their emotions experienced improvements in their psychological well-being. The Linguistic Inquiry and Word Count (LIWC) was created because humans were not efficient or accurate enough when reading texts (LIWC 2015 dictionary categories can be seen in Table A1) [17]. The LIWC calculates psychologically meaningful words to determine a person's psychological state. The dictionary has been widely used in linguistic analyses on topics such as depression, intimacy, and group psychology research [18]. A Chinese Linguistic Inquiry and Word Count (CLIWC) tool was developed by Taiwanese scholars, based on the unique categories of the Chinese language. The CLIWC has established reliability and validity, and can be used to detect mental health issues in Chinese texts with high accuracy [19].

Based on the psychological diagnosis results of linguistic analyses, an online suicide prevention and psychological intervention system could be established for LBC. Fang, Chang, and Fan extracted characteristic words related to mental disorders through text matching of depressive dairies published on social media, and then scored them [20]. When the score exceeded a certain level, the system would automatically send messages to the patient's friends, alerting them that the patient might need psychological support. Parime and Suri used machine learning techniques to monitor online bullying on social media, in order to develop interventions that could reduce bullying [21]. Yang et al. used artificial intelligence technology to collect suicide-related keywords, such as "jumping off a building, cutting wrists, burning charcoal, and jumping into a river" to identify Weibo users with suicidal ideations, therefore helping to establish a suicide intervention mechanism [22]. Moreover, existing studies have mainly focused on depression, bipolar disorder, cyber bullying, and other groups, but few studies have focused on LBC. In this study, we identified LBC's psychological problems through linguistic analysis and text mining by detected LBC with suicidal ideations, through their use of death-related words. The findings can be used to establish an online suicide prevention or psychological intervention for LBC in the future.

2. Methods

2.1. Data Source

In contrast to traditional data collection methods, such as questionnaires and standardized scales, in this study we created a computing application to quantitatively collect personal narrative texts posted by LBC on Zhihu, a Chinese question-and-answer community website. Zhihu is similar to Quora in the United States, with two functions; namely, social networking and question-and-answer (Q&A) [23]. As of November 2018, Zhihu had more than 220 million users, posting more than 30 million questions and more than 130 million answers [24]. As China's largest user content production platform, Zhihu contains user information on various groups in its massive text answers [25].

In this study, we used personal narrative texts under the three most popular questions (according to the views) posted about LBC on the Zhihu website as the research text data, with a total of 1475 personal narrative texts (1,013,829 words). Table 1 list the three most popular questions posted about LBC on the Zhihu.

Table 1. The three most popular questions posted about LBC on the Zhihu website.

Questions about LBC on Zhihu
Q_1 What is the status of left-behind children when they grow up?
Q_2 What is the experience of left-behind children?
Q_3 What kind of psychological problems do left-behind children have when they grow up?

Regarding each of the three questions, the occupation of the questioner, time the question was posted, and degree of attention it received are summarized in Table 2. In order to comply with the data privacy rights of the questioner, the ID name is represented by a code.

Table 2. Basic information on the three studied questions posted on the Zhihu website.

Question	Questioner ID	Questioner’s Occupation	Time	Degree of Attention	
				Question Followers	Number of Views
Q_1	ID1	Art worker	25 July 2014	3936	1,403,306
Q_2	ID2	Parenting teacher	3 November 2015	1627	712,961
Q_3	Anonymous	Unknown	10 May 2014	551	334,302

The details of the three questions about LBC are shown in Table 3, including the number of answers, the time frame for answering the questions, and the gender of the responders.

Table 3. Basic information on responses to the three questions about LBC.

Question	Number of Responses	Time Frame of Posted Answers	Gender of Responder		
			Male	Female	Unknown
Q_1	776	July 2014~December 2019	542	144	80
Q_2	561	November 2015~December 2019	345	137	79
Q_3	138	May 2015~January 2020	90	24	24

2.2. Data Collection

In this study, a web crawler computing application was implemented to gather first-hand textual data by accessing the webpage data, through the designed code of computer programming languages R and Python, to obtain useful data from massive data [26,27]. The process was implemented using Python programming, as shown in Figure 1. Selenium WebDriver technology was used to crawl Zhihu webpages and obtain all LBC’s personal narrative texts, as mentioned above.

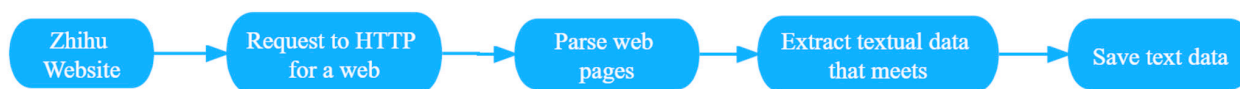


Figure 1. Basic principles of a web crawler.

To avoid violating the privacy protection of social media users’ personal information, a user’s ID number and name must be desensitized [28]. Thus, the application only collected the text content, time, and background information of the Q&A. The text data collected above were retrieved on 15 April 2020, and were analyzed for the first time.

2.3. Data Analysis

There are multiple studies that have performed data analyses from various positions, such as specialists in programming, policy, and child psychology. The analyses are produced through the same, or very similar, interpretations of researchers, to avoid the possibility of bias and reassess the assumptions, as well as to better determine a point of view which shows the most objectivity overall. The research steps used in this study are described below.

2.3.1. Linguistic Analysis

Because the text of this study was in Chinese, the latest version of the CLIWC dictionary (from 2015), which includes 9720 types of Chinese vocabulary, was used. As early as 40 years ago, physician Walter Weintraub found that the number of times a person used first-person singular pronouns was related to people’s depression [18]. Since then, other scholars have found that people with depression or suicidal ideations used more first-person singular pronouns, negative words, past tense verbs, and death-related words in their writing [29,30]. Based on previous studies, we hypothesized that LBC would

use negative emotion-based words, first-person singular pronouns, past tense verbs, and death-related words, which, in turn, relate to a higher degree of depression and stronger suicidal ideations. Therefore, four types of vocabulary in the CLIWC dictionary were adopted in this study: first-person singular pronouns, negative words, past tense verbs, and death-related words. Then, 1475 personal narrative texts of LBC on Zhihu (featuring 1,013,829 words), and a control group, were analyzed. If the proportions of the four types of vocabulary in the LBC text were higher than those of the control group, bilateral *t*-tests were used to test statistical differences between the two groups. The steps of the analysis are as follows:

- a. Propose the research hypothesis.

Hypothesis 0 (H0). *In the personal narrative text of LBC and the control group, the proportion of the four types of vocabulary in the total number of words is the same.*

Hypothesis 1 (H1). *In the personal narrative text of LBC and the control group, the proportion of the four types of vocabulary in the total number of words is different.*

- b. Calculate the statistics.

The two samples of LBC and control group texts have a known sample mean and standard deviation, and approximate a normal distribution. Therefore, in this study, we used a bilateral *t*-test to determine whether the two-sample means were significantly different [31]. The equation of the independent-sample *t*-test was as follows [32]:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \quad (1)$$

where \bar{X}_1 and \bar{X}_2 are the means of the two samples, respectively; n_1 and n_2 are the sizes of the two samples, respectively; and S_1^2 and S_2^2 are the variance of the two samples, respectively.

- c. Determine whether statistically significant differences exist between the study group and the control group.

First, we determined the *p*-value by checking the critical value table. If $p < 0.05$, the type of vocabulary in the text of the LBC group and the control group is statistically significant, rejecting H_0 .

All algorithms and results for the linguistic analysis were implemented by the research team in Python programming.

2.3.2. Textual Analysis

A textual analysis obtains information by understanding the language in a body of text [33]. It is a generalized qualitative analysis method and a mainstream form of research in social science [34]. It does not emphasize the calculation and quantitative analysis of text, but uncovers hidden viewpoints and truth through theory, thereby effectively resolving the oppression that marginalized groups face in traditional research design [35,36].

To present a picture of LBC's experiences, as well as to explore their psychological statuses and contributing factors, in this study we retrieved representative LBC's texts through a lexical matching-based model [37]. The key words used for data extraction are summarized in Table 4.

Table 4. Key words for data extraction.

Lexical for Searching LBC's Text	
Negative words	Get out, nightmares, quarrel, bullying, sexual assault, self-abasement, depression.
Death-related words	Suicide, taking drugs, die, despair, cut wrist.

3. Results

3.1. The Linguistic Analysis of LBC

To examine the mental state of the LBC in this study, we counted four types of vocabulary associated with depression and suicide in the texts of LBC, namely first-person singular pronouns, negative words, past tense verbs, and words associated with death, by text mining, and then calculated the proportion of total words for each type of vocabulary. Then, the personal narrative texts under the topic, but unrelated to LBC, on Zhihu, were randomly extracted as the control group. The specific text data can be found in Table A2. The counting results can be seen in Table 5, and the histogram can be seen in Figure 2.

Table 5. Count results of four types of vocabulary between the LBC group and the control group.

Four Types of Vocabulary	LBC		Control Group	
	Counts	Proportion of Total Words	Counts	Proportion of Total Words
First-person singular pronoun	39,723	3.92%	1530	0.73%
Negative words	46,655	4.60%	3212	1.54%
Past tense verbs	12,351	1.22%	959	0.46%
Death-related words	3049	0.30%	311	0.15%
Total	101,778	10.40%	6012	2.88%

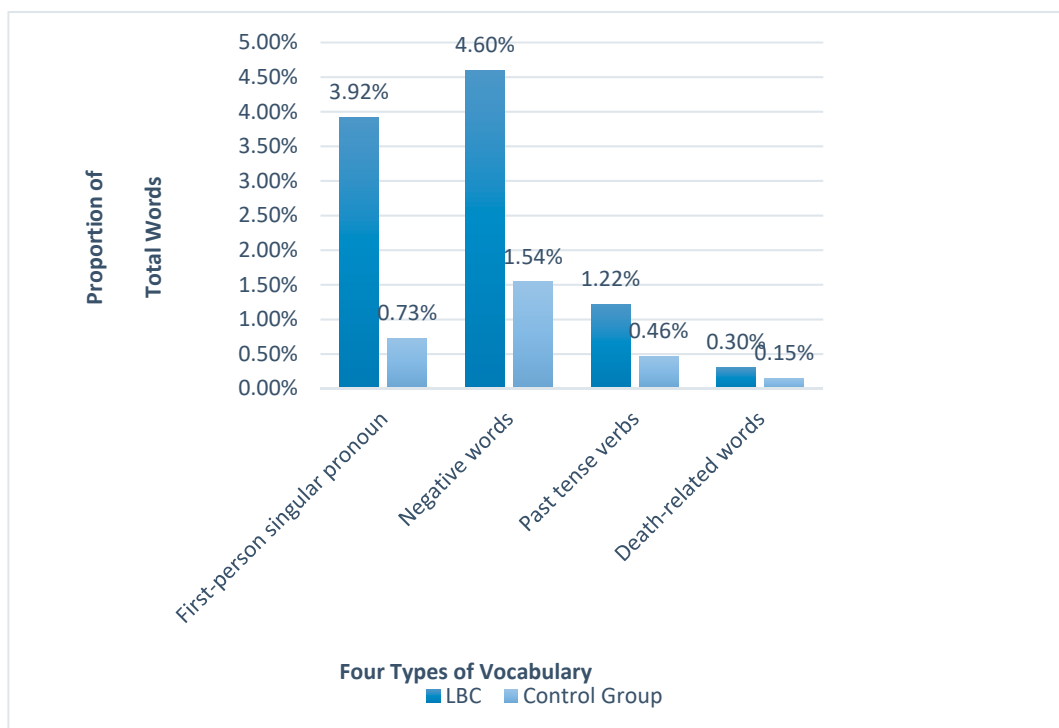


Figure 2. Histogram of four types of vocabulary between the LBC group and the control group.

As can be seen in Table 5, LBC used the four types of vocabulary more than the control group. The proportion of first-person singular pronouns among LBC was 5.37 times that of the control group, followed by negative words (2.99 times), past tense verbs (2.65 times), and death-related words (2.00 times). On this basis, bilateral *t*-tests were performed on the two sets of texts to determine whether the differences were statistically significant. Table 6 shows that the proportions of the four types of vocabulary used between the LBC and control groups were significant.

Table 6. Bilateral *t*-test results of personal narrative text differences between the LBC group and the control group.

	First-Person Singular Pronouns	Negative Words	Past Tense Verbs	Death-Related Words
<i>t</i>	−20.0929	−15.8535	−14.0989	−9.65936
<i>p</i> -Value	2.76×10^{-81}	5.90×10^{-52}	2.19×10^{-42}	3.36×10^{-21}

$p < 0.05$.

3.2. The Textual Analysis of LBC

Following the linguistic analysis of personal narrative texts, we extracted five representative personal narrative texts of LBC related to child neglect, and five representative personal narrative texts of LBC related to depression and suicide, as shown in Tables 7 and 8.

Table 7. Five representative personal narrative texts of LBC related to child neglect.

LBC_1: Every time I call [my parents], their first sentence will always be: “What happened?” I said nothing. “I just wanted to talk to you.” The rest of the conversation would be, they would answer whatever I asked, and then there would be no follow-up. Now when I call them, it’s mostly because I went out of money. It’s not often, once a month. It’s like having a period, which sometimes it’s early but sometimes late.

LBC_2: What I heard most since I was a child is, “Your parents work hard for you to make money, so you have to be obedient and don’t disappoint your parents.” “Your parents don’t want you anymore, so you need to live in someone else’s house.” “Get out of here, who wants your parents’ rotten money?” “Your parents have done a lot of evil in their previous life to have a child like you, just like a debt collector.” “If you hadn’t been born, your brother would have been the only child. How happy your parents would be.” . . . It’s really like a nightmare.

LBC_3: Although my parents sent letters and made phone calls from time to time, they never sent me a penny. All I remember is that my grandmother had a bad temper and used to quarrel and fight with my grandfather. Every day I took a lunch box and rice to the school dining room with steamed rice, and then I took some pickles and ate them. Occasionally, my grandmother gave me a dime to buy food at school (I vaguely remember the food provided by the school is 1-3 dimes). The child does not understand the hardships of life at all (my grandmother recalled that I hadn’t eaten any sugar for three years). Then such days continued until nearly 6th grade.

LBC_4: When I was bullied at school, other children had parents supporting them, but I didn’t, so I could only endure sexual abuse. I had nowhere to ask for help when I was sexually assaulted (let me be anonymous). I am good at studying. My relatives like me and my teachers also like me, but I am still an introverted and self-abasement person.

LBC_5: There are still many drawbacks since I was little without my parents around me. For example, I haven’t developed good study habits since I was young. There was very little homework in the countryside, so I could get first place casually in primary school. I have never studied seriously, and my grades were very unstable when I went to school in the city. Before I went to school in the city, I never knew that I needed to take notes in class, because no one took notes in my countryside school, and the teachers would not tell you to do so as well. I never studied English in the rural elementary school for six years (there are English classes, but you know that).

Table 8. Five representative personal narrative texts of LBC related to depression and suicide.

LBC_6: When I first entered college, I felt low self-esteem so that I shed tears in the middle of the night. I like to put pressure on myself, hoping to make myself better. I don’t care about people in relationships and don’t know how to socialize with opposite-gender friends. Apart from genetic inheritance, the main reason for that is the experience of left-behind children. No matter what I can achieve in the future, I will not be happy inside.

LBC_7: Now I am inferior, sensitive, and insecure. I feel that my parents are terribly unfamiliar, and sometimes I feel embarrassed to stay with them. I hated them, why they gave birth to us but don’t care about us.

LBC_8: As a senior left-behind child, I lack of love, insecure, strong defensive heart, not talkative, and love to be alone.

LBC_9: I had severe depression five years ago. At that time, I had auditory hallucinations and committed suicide twice, and I am still taking medication. I was rescued by my parents. The first time I used a pair of tweezers in the toilet to stab my chest, and the second time I used underwear to strangle my neck in the hospital. Later, my mother called 120 (emergency) for me. I don’t know if everyone who wants to die has such despair as me. It just feels that in this world, there is really no love and no meaning in life.

LBC_10: Hehe. I hate being with them, hate their words, hate their selfishness, hate their ignorance. At some point, I wish I had died suddenly.

4. Discussion

4.1. Risk of Depression and Suicide

The present findings show that the proportions of four types of vocabulary (first-person singular pronouns, negative words, past tense verbs, and death-related words) in LBC personal narrative texts were all significantly higher than those in the control group, especially the proportion of first-person singular pronouns (5.37 times as high). Brockmeyer et al. reported that frequent use of first-person singular pronouns was associated with self-focused attention, which is a cognitive bias highly related to depression [38]. In a meta-analysis study, Holtzman argued that the use of first-person singular pronouns is a linguistic feature of people with depression, and that there is a correlation between the two [39]. Increased use of first-person singular pronouns has been linked to social isolation and suicide [40]. Therefore, the high rate of first-person singular pronouns among LBC suggests obvious self-focused characteristics, an important sign of depression and suicidal ideations.

The rate of use by LBC of negative words was 2.99 times that of the control group. Studies have shown that people with anxiety and depression showed attentional biases toward negative words [41]. People with depression showed less pleasure and arousal relative to positive words, but increased arousal relative to negative words [42]. The proportion of LBC using past tense verbs was 2.65 times higher than the control group, consistent with the finding that people with depression pay more attention to past events [43]. This suggests an explanation for why depression is often regarded as “being trapped in the past”, and is also reflected in the high use of past tense verbs in written disclosures [44,45].

With respect to death-related words, LBC used these words at twice the rate of the control group. In an analysis of letters written in the 2 years before suicide, researchers found that the frequency of using death-related words increased significantly [46]. Based on the linguistic analysis results, negative words, past tense verbs, and words associated with death in the texts of LBC were significantly more prevalent than those of the control group.

The findings of this study showed that LBC tended to use more first-person singular pronouns, negative words, past tense verbs, and death-related words. This suggests that LBC are deeply stuck in their painful left-behind experience, and excessively focus on themselves and their pasts. It is even more alarming that they are emotionally negative, and that some of them have attempted suicide or have suicidal thoughts and ideations.

4.2. Child Neglect as a Source of Depression and Suicidal Thoughts among the LBC in this Study

As previously discussed, the LBC in this study showed significant characteristics of depression and suicidal ideations. This research incorporated textual analysis of LBC personal narrative texts to present a deeper picture of possible causes.

Recent research has indicated that the proportion and degree of neglect experienced by LBC were higher than that of other children, especially, in rural areas in Western China [47]. LBC can be more prone to mental trauma and traumatic experiences than their counterparts [48]. Analyzing LBC’s personal narrative texts indicated that their mental well-being could be traced back to various types of neglect in early childhood.

Child neglect is defined as not meeting the basic needs of children, including adequate supervision, health care, clothing, and housing, as well as other physical, emotional, social, educational, and safety needs [49]. Child neglect is the most common type of child maltreatment, and affects the development and well-being of children worldwide [50]. However, because the damage caused by neglect is difficult to detect, and quantitatively measure, it has received far less attention in the field than child abuse [51].

First, the type of neglect that is unavoidable for LBC is parental emotional neglect. LBC are separated from at least one of their parents during childhood. Parental emotional neglect is related to externalizing problem behaviors [52], and is also a major risk factor for psychopathology, including internal problems such as depression [53]. Second, LBC’s parents migrate from underdeveloped areas to developed cities and participate in China’s

economic growth. Due to the limited income conditions of migrant workers, many of them cannot provide a promising life for their LBC. This contributes to a certain degree of neglect related to health care, clothing, food safety, education, social, and other aspects of LBC. For example, the incidence of infectious and chronic diseases among rural LBC is 20% higher than that of non-LBC [54]. Their sacrifices are also seen as the price of China's economic prosperity [55]. Finally, there is a lack of adequate supervision in addition to physical, social, and educational neglect. Among LBC in rural China, the percentages of LBC who experience victimization and polyvictimization have been reported to be 27.5% and 10.94%, respectively, and the poverty of LBC has been positively correlated to polyvictimization [56]. In addition, due to the lack of caregiver supervision, and the closed and monotonous living environment in LBC's hometown, they often experience social and educational neglect. Some of them are also subjected to physical abuse and sexual assault [57]. These findings clearly show that LBC encounter various types and levels of child neglect during childhood, which lead to a higher risk of depression, suicide [58], and psychological trauma [59].

4.3. Online Suicide Prevention and Positive Intervention for LBC

The results of this study indicate that more LBC have severe experiences of child neglect, and therefore are more prone to depression and suicide. As illustrated in Table 7, some individuals with left-behind experience have suicidal ideations, or have exhibited suicidal behaviors. Suicide is a serious global public health problem [60]. The latest number of left-behind children in China reached 290 million [2]; therefore, it is imperative to establish an online suicide prevention system for LBC.

Social media is one of the important channels for vulnerable groups to express their emotions, and research has shown that they are more willing to self-expose in a public channel than in a private space [22]. Wang, et al. indicated that self-presentation in an online environment contributed to the treatment of depression [61]. Online technology is also increasingly being used in health and psychological care [62]; therefore, an online LBC group could be established to share and discuss their left-behind experiences and provide support for each other. In addition, spiritual leaders could share their successful experiences to help other helpless LBC cope with depression, and to teach them how to develop the ability to think positively. Meanwhile, digital technology has enabled a new way of thinking; i.e., that services and treatment could be provided online [63]. For example, social workers could offer positive psychological interventions through Zhihu's LBC community. Therefore, it is conceivable to establish an online psychological intervention system for LBC. The relevant public departments could consider digital transformation of social work practices to innovatively solve the country's public health crisis. This exploratory study demonstrates an opportunity for digital help for LBC, and offers new solutions to LBC's public health problems.

5. Conclusions

This study used the psychometric dictionary CLIWC to conduct text mining through linguistic analysis, combined with textual analysis, of personal narrative texts written by LBC, which were collected from a social media website, Zhihu, in mainland China. Three key findings, obtained via quantitative analysis, are presented for discussion. First, LBC show a higher risk of depression and suicidal ideations than other children. Second, the psychological trauma of LBC stems from the different types and degrees of neglect suffered in their childhood. Finally, text mining was utilized to extract death-related words to identify LBC with suicidal risk, making a case to establish an online suicide prevention and psychological intervention system. More importantly, it provides a forward-looking approach that can be used for future research on the mental health and well-being of LBC and other vulnerable populations.

Inevitably, there are some limitations to this study, primarily concerning how to determine the authenticity of content on social media. It would be beneficial to consider a mixed-methods design; for example, combining big data analytics with traditional data

collection methods to ensure the quality, rigor, and authenticity of the data. Text mining technology could be used as a supplement and extension of traditional research methods. For example, researchers could select social media users who fit the research topic and invite them to conduct online focus groups. From the perspective of digital public health, text mining technology could be used as a digital technology to improve public health problems. In summary, the exploratory research work of this study has important practical significance. It should facilitate the practical application of digital technologies in addressing public health problems.

Author Contributions: Conceptualization, Y.L. and J.C.-C.C.; methodology, Z.L., C.R. and Y.L.; software, Z.L.; validation, Y.L.; formal analysis, Y.L.; investigation, Y.L.; resources, Y.L.; data curation, Y.L., Z.L. and J.X.; writing—original draft preparation, Y.L.; writing—review and editing, J.C.-C.C. and J.-J.H.; visualization, Y.L.; supervision, J.C.-C.C., J.X. and J.-J.H.; project administration J.C.-C.C. and J.X.; funding acquisition, J.X. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the China Scholarship Council (grant number 201808440484).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Classification of function words in LIWC Dictionary.

Total Function	Pronouns	Other Function	Other Grammar	Affect	Social	Cognitive Processes	Perceptual Processes	Relativity	Personal Concerns	Informal Language
	I	Articles	Verbs	Positive	Family	Insight	See	Motion	Work	Swear
	We	Prepositions	Adjectives	Negative	Friend	Causation	Hear	Space	Leisure	Net speak
	You	Auxiliary verbs	Comparisons	Anxiety	Female	Discrepancy	Feel	Time	Home	Agreement
	She and he	Adverbs	Interrogatives	Anger	Male	Tentative			Money	Non-fluencies
	They	Conjunctions	Numbers	Sadness		Certainty			Religion	Filler
	Impersonal	Negations	Quantifiers			Difference			Death	

Table A2. Text information of control group.

Questions of Control Group	Number of Answer
How do you spend your day?	297
What’s your daily routine like?	66
Total	363
Total number of words	209,119

References

1. Fang, C.; Yang, D.; Meiyang, W. *Migration and Labor Mobility in China*; Munich Personal RePEc Archive: Munich, Germany, 2009.
2. National Bureau of Statistics. *The Main Data Bulletin of the National 1% Population Sample Survey in 2015*; National Bureau of Statistics: Beijing, China, 2016.
3. Duan, C.R.; Lyu, L.D.; Zou, X.J. Major challenges for China’s floating population and policy suggestions: An analysis of the 2010 Population Census Data. *Popul. Res.* **2013**, *37*, 17–24.
4. Duan, C.R.; Zhou, F.L. Research on the Situation of Left-Behind Children in China. *Popul. Res.* **2005**, *1*, 29–36.

5. State Council Information Office. National Family Data Release: Rural Left-Behind Children Account for More Than 1/3. Available online: <http://www.scio.gov.cn/zhzc/8/4/Document/1433950/1433950.htm> (accessed on 22 January 2022).
6. Ministry of Civil Affairs. Chart: 2018 Rural Left-Behind Children Data. Available online: <http://www.mca.gov.cn/article/gk/tjtb/201809/20180900010882.shtml> (accessed on 2 February 2022).
7. Federation, Research Group of All-China Women's. The All-China Women's Federation Issued a Report on the Situation of Left-Behind Children in Rural Areas and Migrant Children in Urban and Rural Areas. Available online: <https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFD2013&filename=ZFYZ201306009&v=9qJhL70VmyWdO5LBl6qT4rKvTaj7nIvImRBq2hPYUwDyHW1sfv0fJSZh%25mmd2FDhMLRz1> (accessed on 25 January 2022).
8. Gao, Y.; Li, L.P.; Kim, J.H.; Congdon, N.; Lau, J.; Griffiths, S. The impact of parental migration on health status and health behaviours among left behind adolescent school children in China. *BMC Public Health* **2010**, *10*, 56. [CrossRef]
9. Jia, Z.; Tian, W. Loneliness of Left-Behind Children: A Cross-Sectional Survey in a Sample of Rural China. *Child. Care. Health Dev.* **2010**, *36*, 812–817. [CrossRef]
10. Guo, J.; Chen, L.; Wang, X.; Liu, Y.; Chui, C.H.K.; He, H.; Qu, Z.; Tian, D. The relationship between internet addiction and depression among migrant children and left-behind children in China. *Cyberpsychol. Behav. Soc. Netw.* **2012**, *15*, 585–590. [CrossRef]
11. Hu, H.; Lu, S.; Huang, C.C. The psychological and behavioral outcomes of migrant and left-behind children in China. *Child. Youth Serv. Rev.* **2014**, *46*, 1–10. [CrossRef]
12. Chen, E.E.; Wojcik, S.P. A practical guide to big data research in psychology. *Psychol. Methods* **2016**, *21*, 458–474. [CrossRef]
13. Aggarwal, C.C.; Zhai, C. (Eds.) An introduction to text mining. In *Mining Text Data*; Springer: Berlin/Heidelberg, Germany, 2012. [CrossRef]
14. Lyu, Y.; Chow, J.C.C.; Hwang, J.J. Exploring public attitudes of child abuse in mainland China: A sentiment analysis of China's social media Weibo. *Child. Youth Serv. Rev.* **2020**, *116*, 105250. [CrossRef]
15. Li, S.; Lu, S.; Ni, S.; Peng, K. Identifying psychological resilience in Chinese migrant youth through multidisciplinary language pattern decoding. *Child. Youth Serv. Rev.* **2019**, *107*, 104506. [CrossRef]
16. Gottschalk, L.A.; Gleser, G.C. *The Measurement of Psychological States through the Content Analysis of Verbal Behavior*; University of California Press: Berkeley, CA, USA, 1979.
17. Pennebaker, J.W.; Boyd, R.L.; Jordan, K.; Blackburn, K. *The Development and Psychometric Properties of Liwc2015*; University of Texas at Austin: Austin, TX, USA, 2015.
18. Tausczik, Y.R.; Pennebaker, J.W. The psychological meaning of words: LIWC and computerized text analysis methods. *J. Lang. Soc. Psychol.* **2010**, *29*, 24–54. [CrossRef]
19. Huang, C.-L.; Chung, C.K.; Hui, N.; Lin, Y.-C.; Seih, Y.-T.; Lam, B.C.P.; Chen, W.-C.; Bond, M.H.; Pennebaker, J.W. The development of the Chinese linguistic inquiry and word count dictionary. *Chin. J. Psychol.* **2012**, *54*, 185–201.
20. Fang, Y.E.; Tai, C.H.; Chang, Y.S.; Fan, C.T. A mental disorder early warning approach by observing depression symptom in social diary. In Proceedings of the 2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC), San Diego, CA, USA, 5–8 October 2014; pp. 2060–2065. [CrossRef]
21. Parime, S.; Suri, V. Cyberbullying detection and prevention: Data mining and psychological perspective. In Proceedings of the 2014 International Conference on Circuits, Power and Computing Technologies [ICCPCT-2014], Nagercoil, India, 20–21 March 2014; pp. 1541–1547. [CrossRef]
22. Yang, D.; Yao, Z.; Seering, J.; Kraut, R. The channel matters: Self-disclosure, reciprocity and social support in online cancer support groups. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, UK, 4–9 May 2019; pp. 1–15. [CrossRef]
23. Chen, J.; Deng, S. An empirical analysis of the influencing factors of user experience on social Q&A platforms—taking Zhihu as an example. *J. Libr. Inf. Serv.* **2015**, *59*, 102–108.
24. Zhihu Product Analysis Report." Beijing, China: SOHU, 2018. Available online: https://www.sohu.com/a/281528441_485557 (accessed on 8 February 2022).
25. Liu, C.J.; Yang, G.; Zhang, R.Y. Research on the publishing transformation of social question-and-answer communities—taking Zhihu as an example. *J. Sci. Technol. Publ.* **2017**, *36*, 102–108.
26. Hejing, W.; Fang, L.; Long, Z.; Yabin, S.; Ran, C. Application research of crawler and data analysis based on Python. *Assoc. Ed. -in-Chief* **2020**, 64–70. [CrossRef]
27. Thomas, D.M.; Mathur, S. Data analysis by web scraping using python. In Proceedings of the 2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 12–14 June 2019; pp. 450–454. [CrossRef]
28. Sokolova, M.; Matwin, S. Personal privacy protection in time of big data. In *Challenges in Computational Statistics and Data Mining*; Springer: Cham, Switzerland, 2016; pp. 365–380. [CrossRef]
29. O'dea, B.; Larsen, M.E.; Batterham, P.J.; Calear, A.L.; Christensen, H. A linguistic analysis of suicide-related Twitter posts. *Crisis* **2017**, *38*, 319–329. [CrossRef]
30. Olsson, V.; Lindow, M. How Does Bipolar and Depressive Diagnoses Reflect in Linguistic Usage on Twitter: A Study Using LIWC and Other Tools. 2018. Available online: <http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-229439> (accessed on 28 January 2022).

31. Jia, J.P. *Statistics*; Tsinghua University Press: Beijing, China, 2006. Available online: https://books.google.com.hk/books?hl=en&lr=&id=KFgJIs7m4YAC&oi=fnd&pg=PA5&dq=%E7%BB%9F%E8%AE%A1%E5%AD%A6&ots=47Y4soL5TS&sig=1zF-aeYTPNdil1WobSuqvw-qDgs&redir_esc=y#v=onepage&q=%E7%BB%9F%E8%AE%A1%E5%AD%A6&f=false (accessed on 9 January 2022).
32. Cressie, N.A.C.; Whitford, H.J. How to Use the Two Sample T-Test. *Biom. J.* **1986**, *28*, 131–148. [[CrossRef](#)]
33. McKee, A. *Textual Analysis: A Beginner's Guide*; Sage: Thousand Oaks, CA, USA, 2003.
34. Fairclough, N. *Analysing Discourse: Textual Analysis for Social Research*; Psychology Press: East Sussex, UK, 2003.
35. Allen, M. (Ed.) *The Sage Encyclopedia of Communication Research Methods*; Sage: Thousand Oaks, CA, USA, 2017. [[CrossRef](#)]
36. Baxter, J. Text and textual analysis. In *International Encyclopedia of Human Geography*; Kitchin, R., Thrift, N., Eds.; Elsevier: Amsterdam, The Netherlands, 2020; pp. 239–243.
37. Tran, V.; Nguyen, M.L.; Satoh, K. Building legal case retrieval systems with lexical matching and summarization using a pre-trained phrase scoring model. In Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law, Montreal, QC, Canada, 17–21 June 2019; pp. 275–282. [[CrossRef](#)]
38. Brockmeyer, T.; Zimmermann, J.; Kulesa, D.; Hautzinger, M.; Bents, H.; Friederich, H.C.; Herzog, W.; Backenstrass, M. Me, myself, and I: Self-referent word use as an indicator of self-focused attention in relation to depression and anxiety. *Front. Psychol.* **2015**, *6*, 1564. [[CrossRef](#)]
39. Holtzman, N.S. A meta-analysis of correlations between depression and first person singular pronoun use. *J. Res. Personal.* **2017**, *68*, 63–68. [[CrossRef](#)]
40. Li, T.M.; Chau, M.; Yip, P.S.; Wong, P.W. Temporal and computerized psycholinguistic analysis of the blog of a Chinese adolescent suicide. *Crisis* **2014**, *35*, 168–175. [[CrossRef](#)]
41. Mogg, K.; Bradley, B.P.; Williams, R. Attentional bias in anxiety and depression: The role of awareness. *Br. J. Clin. Psychol.* **1995**, *34*, 17–36. [[CrossRef](#)]
42. Liu, W.H.; Wang, L.Z.; Zhao, S.H.; Ning, Y.P.; Chan, R.C. Anhedonia and emotional word memory in patients with depression. *Psychiatry Res.* **2012**, *200*, 361–367. [[CrossRef](#)]
43. Eysenck, M.; Payne, S.; Santos, R. Anxiety and depression: Past, present, and future events. *Cogn. Emot.* **2006**, *20*, 274–294. [[CrossRef](#)]
44. Holman, E.A.; Silver, R.C. Getting “stuck” in the past: Temporal orientation and coping with trauma. *J. Personal. Soc. Psychol.* **1998**, *74*, 1146. [[CrossRef](#)]
45. Rodriguez, A.J.; Holleran, S.E.; Mehl, M.R. Reading between the Lines: The Lay Assessment of Subclinical Depression from Written Self-Descriptions. *J. Personal.* **2010**, *78*, 575–598. [[CrossRef](#)]
46. Barnes, D.H.; Lawal-Solarin, F.W.; Lester, D. Letters from a suicide. *Death Stud.* **2007**, *31*, 671–678. [[CrossRef](#)]
47. Zhong, Y.; Zhong, Z.H.; Pan, J.P.; Wang, Y.X.; Liu, C.Y.; Yang, X.; Hu, C.; Cai, L.L.; Xu, Y. The situation of children neglect between left-behind children and living-with-parents children in rural areas of two western provinces of China. *Chin. J. Prev. Med.* **2012**, *46*, 38–41.
48. Sun, M.; Xue, Z.; Zhang, W.; Guo, R.; Hu, A.; Li, Y.; Mwansisya, T.E.; Zhou, L.; Liu, C.; Chen, X.; et al. Psychotic-like experiences, trauma and related risk factors among “left-behind” children in China. *Schizophr. Res.* **2017**, *181*, 43–48. [[CrossRef](#)]
49. Bovarnick, S. *Child Neglect: NSPCC Child Protection Research Briefing*; NSPCC: London, UK, 2007.
50. Hildyard, K.L.; Wolfe, D.A. Child neglect: Developmental issues and outcomes. *Child Abuse. Negl.* **2002**, *26*, 679–695. [[CrossRef](#)]
51. Garbarino, J.; Collins, C.C. Child neglect: The family with a hole in the middle. In *Neglected Children: Research, Practice, and Policy*; Dubowitz, H., Ed.; Sage: Thousand Oaks, CA, USA, 1999; pp. 1–23.
52. Yang, B.; Xiong, C.; Huang, J. Parental emotional neglect and left-behind children’s externalizing problem behaviors: The mediating role of deviant peer affiliation and the moderating role of beliefs about adversity. *Child. Youth Serv. Rev.* **2021**, *120*, 105710. [[CrossRef](#)]
53. Young, R.; Lennie, S.; Minnis, H. Children’s perceptions of parental emotional neglect and control and psychopathology. *J. Child Psychol. Psychiatry* **2011**, *52*, 889–897. [[CrossRef](#)]
54. Li, Q.; Liu, G.; Zang, W. The health of left-behind children in rural China. *China Econ. Rev.* **2015**, *36*, 367–376. [[CrossRef](#)]
55. Jingzhong, Y. Left-behind children: The social price of China’s economic boom. *J. Peasant. Stud.* **2011**, *38*, 613–650. [[CrossRef](#)]
56. Hu, H.; Zhu, X.; Jiang, H.; Li, Y.; Jiang, H.; Zheng, P.; Zhang, C.; Shang, G. The Association and Mediating Mechanism between Poverty and Poly-Victimization of Left-Behind Children in Rural China. *Child. Youth. Serv. Rev.* **2018**, *91*, 22–29. [[CrossRef](#)]
57. Wang, C.; Tang, J.; Liu, T. The sexual abuse and neglect of “left-behind” children in rural China. *J. Child Sex. Abuse.* **2020**, *29*, 586–605. [[CrossRef](#)]
58. Blumberg, M.L. Depression in abused and neglected children. *Am. J. Psychother.* **1981**, *35*, 342–355. [[CrossRef](#)]
59. De Bellis, M.D. The psychobiology of neglect. *Child Maltreatment* **2005**, *10*, 150–172. [[CrossRef](#)]
60. Stone, D.M.; Barber, C.W.; Potter, L. Public health training online: The national center for suicide prevention training. *Am. J. Prev. Med.* **2005**, *29*, 247–251. [[CrossRef](#)]

61. Wang, P.; Wang, X.; Zhao, M.; Wu, Y.; Wang, Y.; Lei, L. Can social networking sites alleviate depression? The relation between authentic online self-presentation and adolescent depression: A mediation model of perceived social support and rumination. *Curr. Psychol.* **2019**, *38*, 1512–1521. [[CrossRef](#)]
62. Best, P.; Manktelow, R.; Taylor, B. Online communication, social media and adolescent wellbeing: A systematic narrative review. *Child. Youth Serv. Rev.* **2014**, *41*, 27–36. [[CrossRef](#)]
63. Reamer, F.G. Clinical Social Work in a Digital Environment: Ethical and Risk-Management Challenges. *Clin. Soc. Work. J.* **2015**, *43*, 120–132. [[CrossRef](#)]