Analyzing the Representation of Body Image in Women's Magazines Through NLP

Tiffany Darmosusilo[1], Bailey Pierson[#] and Julia Deen[#]

[1]Jakarta Intercultural School, Indonesia
[#]Advisor

ABSTRACT

In the age of media and widespread digital presence, there is increasing pressure to conform to unrealistic beauty standards. The media places significant emphasis on body image and appearance-related features, with recent emphasis on the glamorization of a thin body ideal. The idolization of unrealistic body types can have dangerous psychological and physical impacts on women, such as body dysmorphia and disordered eating. This computational study compares and analyzes how body image is portrayed in four online women's magazines — Woman's World, Seventeen, Shape, and Cosmopolitan. In this study, a dataset totaling over 1000 articles was sourced from the online publications of the four magazines. Studying how women's magazines portray body image can be useful in understanding the quantitative measures and raising awareness of the effects of such language. Through the application of Natural Language Processing (NLP) methods of *nltk* relative frequency tasks, we aim to interpret the appearance of body-related language pertaining to thinness, weight loss, and body image. Our results show consistent usage of body-image language for all four magazines. Out of the four magazines, Woman's World has the most significant frequency of language related to weight loss, thinness, and body composition. Both Shape and Woman's World also have prevalent language related to changing body shape.

Background

Body image is defined as the way in which one thinks, perceives, and feels about their body, such as weight, appearance, and body size. Messages and representations of body image are very prevalent in modern media, particularly emphasized in its impact on women. Magazines are one of many media outlets that portray body image and appearance culture, often leading to the popularization and perpetuation of certain body ideals.

The concept of a thin and slender physique is referred to as the thin ideal, which is composed of women with low body weight, little body fat, and very slim feminine features (Aparicio-Martínez et al., 2019). In a study conducted on four women's fashion magazines in the late 1900s, there was a significant decrease in the model's body size and increased body exposure over time (Sypeck et al, 2004).The media is prevalent in representing, and perhaps glamorizing the thin ideal to its large audiences. The representation of thinness in popular culture associates thinness with a celebratory achievement or a motivational necessity to attain happiness and self-satisfaction. One study discovered that certain women's health and fitness magazines glamorized health as thinness rather than accurately portraying healthy body standards (Conlin & Bissell, 2014). In the same study, women's magazines from a wide variety of genres all seemed to place a significant focus on "weight loss for the sake of overall appearance." (Conlin & Bissell, 2014, p.13) The media's overwhelmingly predominant representation and glamorization of thinness reinforces unattainable body ideals to its women audiences. Research shows that individuals with reported high media internalization had consistently high levels of drive for thinness. (Swiatkowski et al, 2016)

Frequent exposure to idealized body image representations or thin body ideals has harmful consequences on a woman's physical, mental, and behavioral aspects (Aparicio-Martínez et al., 2019). The idealization and internalization of a thin female body have devastating effects on body dissatisfaction and body perceptions. Individuals with internalized depictions of body standards in magazines have higher levels of discontent with their bodies and their pursuit of thinness (Swiatkowski, 2016). These statistics suggest that many magazines have a negative impact on body image, due to language or representations of unrealistic body types. Hence, it is important to analyze how these magazines present body-related language and consider the impact of its connotations.

Methodology

Data Collection

For this study, data from online articles from four different women's magazine websites — Woman's World, Seventeen, Cosmopolitan, and Shape — was used.

Table 1. Demographic and Content Information on Four Magazines Used.

Magazine	Age Demographic	Content	Number of Articles in Dataset
Woman's World	Middle aged women over 50 years of age.	General, Health, fitness, beauty, lifestyle.	585
Seventeen	Teenage girls and young women; mean age of 26.	Teen fashion, beauty, entertainment, lifestyle	155
Cosmopolitan	Young women in their 20s and 30s.	fashion, beauty, relationships, sex, and health.	100
Shape	women interested in fitness and health; mean age of 40.	fitness, nutrition, health, and wellness.	300

These magazines were chosen due to their differences in audience demographics and content. Furthermore, all of these magazines have very high readerships — an important consideration regarding the magnitude to which the content of these magazines affects their audiences. With a variety of topics and content focuses being covered by each magazine, there may be different ways and frequencies in which body image is presented, if at all.

To create the dataset, data was collected from the textual data of each magazine. Raw text data was imported from each magazine through web scraping - extracting data from websites and processing it into a usable format. The text data was stored in a corpus and downloaded into separate CSV files for each magazine. Each row of the dataset stored the information of one article.

Each magazine's textual dataset was then preprocessed with custom functions using Python's nltk methods of word tokenization, sentence tokenization, lemmatization, and data cleaning. Tokenization breaks down large chunks of data into discrete units, in order to help the computer interpret its numerical meaning and determine things like word frequency, distribution, etc. Lemmatization is the process of reducing words to their

base or dictionary form. The textual dataset was cleaned to contain only words or sentences, with punctuation, stopwords, in-code lines, HTML tags, hyperlinks, and symbols being removed.

Table 2. Subset of Woman's World Dataset

Magazine	Title	Heading	Category	Text	Date	Author	Tokenized Sentence	Tokenized Words	Individiual Clean Words
Woman's World	whole grain mustard lowers cholesterol, blood sugar	just one teaspoon of whole grain mustard can boost good cholesterol and lower blood sugar …	Diabetes	sometimes, the answer to our biggest health struggles…	July 16 2022	Brenda Kearns	['sometimes, the answer to our biggest health struggles is sitting right in our pantry.' …]	['sometimes', 'the', 'answer', 'to', 'our', 'biggest', 'health', 'struggles', 'is', 'sitting', 'right', 'in', 'our', 'pantry', …]	'sometimes', 'answer', 'biggest', 'health', 'struggles', 'sitting', 'right', 'pantry', 'case', 'whole', 'grain'

Each row of the dataset contained the textual data of the title, heading, category, text, date, author, tokenized sentence, tokenized words, and individual clean words.

To focus on body-image-related language, a list of keywords was created to specifically interpret its frequency and do a content analysis. This list was created from two categories. The first category had keywords pertaining to thinness ('slim', 'skinny', 'fit', 'small', 'bikini-body', 'flat', 'lean', 'small', 'thin', 'toned', 'tight'). The second category had keywords related to body image ('shrink', 'shed', 'slim', 'lose', 'sculpt', 'anorexia', 'disorder', 'bulimia', 'binge-eating', 'lose', 'weight', 'weight loss', 'lose weight', 'eat', 'guilt', 'body shape', 'pounds', 'body', 'frame', 'fat', 'weight', 'ideal', 'diet', 'calories'). The first category examines the representation of a thin body ideal and includes the usage of synonyms pertaining to thinness. The second category explores language often used to promote losing weight or changing one's body to achieve a certain body shape. This set of keywords was used for each task.

The study then created a coding framework that used the textual dataset and performed methods to create a quantitative analysis of the representation of these keywords with each magazine.

Models

Natural Language Processing (NLP) is a domain that uses computers to understand and interpret human language. NLP models use various textual analysis algorithms on data to gain deeper understanding of how human language and computers interact. In this study, Python's *nltk* and *spacy* libraries were used due to their comprehensive tools and efficient models for language processing.

The models used were frequency-based NLP tasks. This study presents the results of seven models:

1. Keyword Search: Searched for the appearance and frequency of these keywords.
2. Temporal Frequency Tasks by Year: How body-image language is used over years. We extracted and concatenated the texts by their respective years. We narrowed down to using the most common words from each year, which we then searched to find the frequency of these body-image words.
3. Temporal Frequency Tasks by Month: How body-image language is used over months. Similarly to Task 2, we extracted and concatenated the texts by their respective month and years. We narrowed down to using the most common words from each month, which we then searched to find the frequency of these body-image words.
4. Percentage comparison of each word for each magazine: We compared the percentage frequency of each body-image word by their magazines.
5. Unique Body-Image Words across Magazines: We found the keywords that were respective to each magazine. We filtered out the most common shared words between all the magazines and found the body-image words.
6. Unique Key-Phrases: Similarly to Task 5, we tokenized the phrases, and found the highest frequency of phrases that were unique to each magazine.
7. Density Comparison of Body Image Phrases: We compared the frequencies of body image phrases for each magazine, and then compared each phrases' densities.

Results

Keyword Search
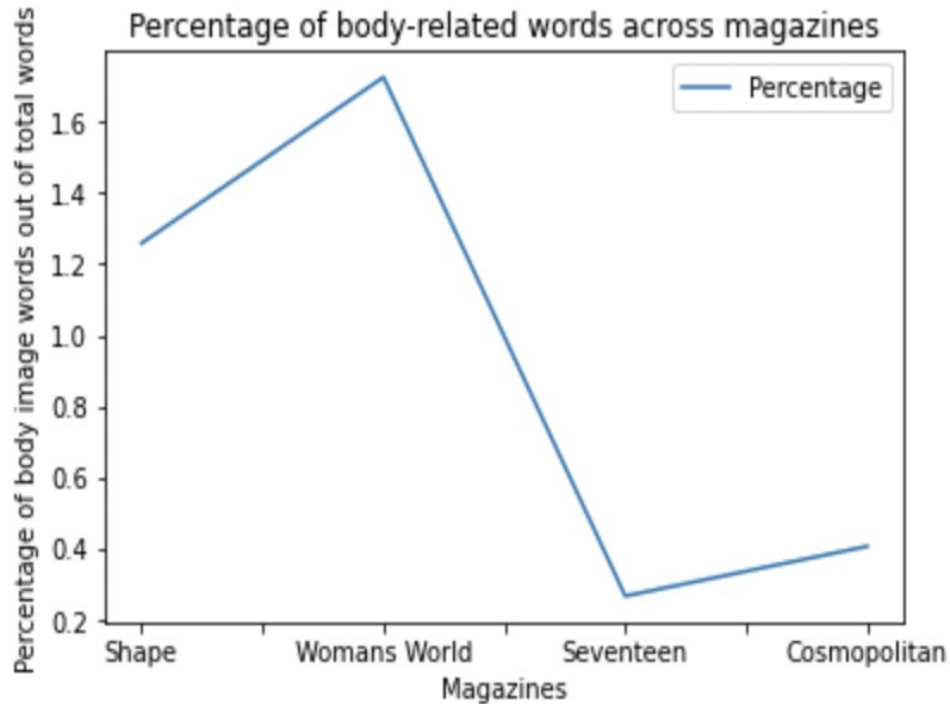
Figure 1. Percentage of body-related words, out of the total word count for each magazine.

Out of the four magazines and their respective word counts, Woman's World has the highest frequency of body-related words, with 1.6% words out of the total magazine's word count pertaining to body-image. Shape has the second highest percentage of around 1.2%. Cosmopolitan and Seventeen have relatively lower frequencies of an estimated 0.4% and 0.25% respectively.

The highest word count from all four magazines was the same word - 'body'. The most frequent words on Woman's World — 'diet', 'fat', 'pounds', 'calories', 'lose', 'weight' — is language largely used to advertise weight-loss methods and encourage diets. This suggests that Woman's World places an emphasis on losing weight. In all the magazines, the frequency of the word 'thin' and thin-related synonyms is very prevalent.

Temporal Frequency Tasks by Year
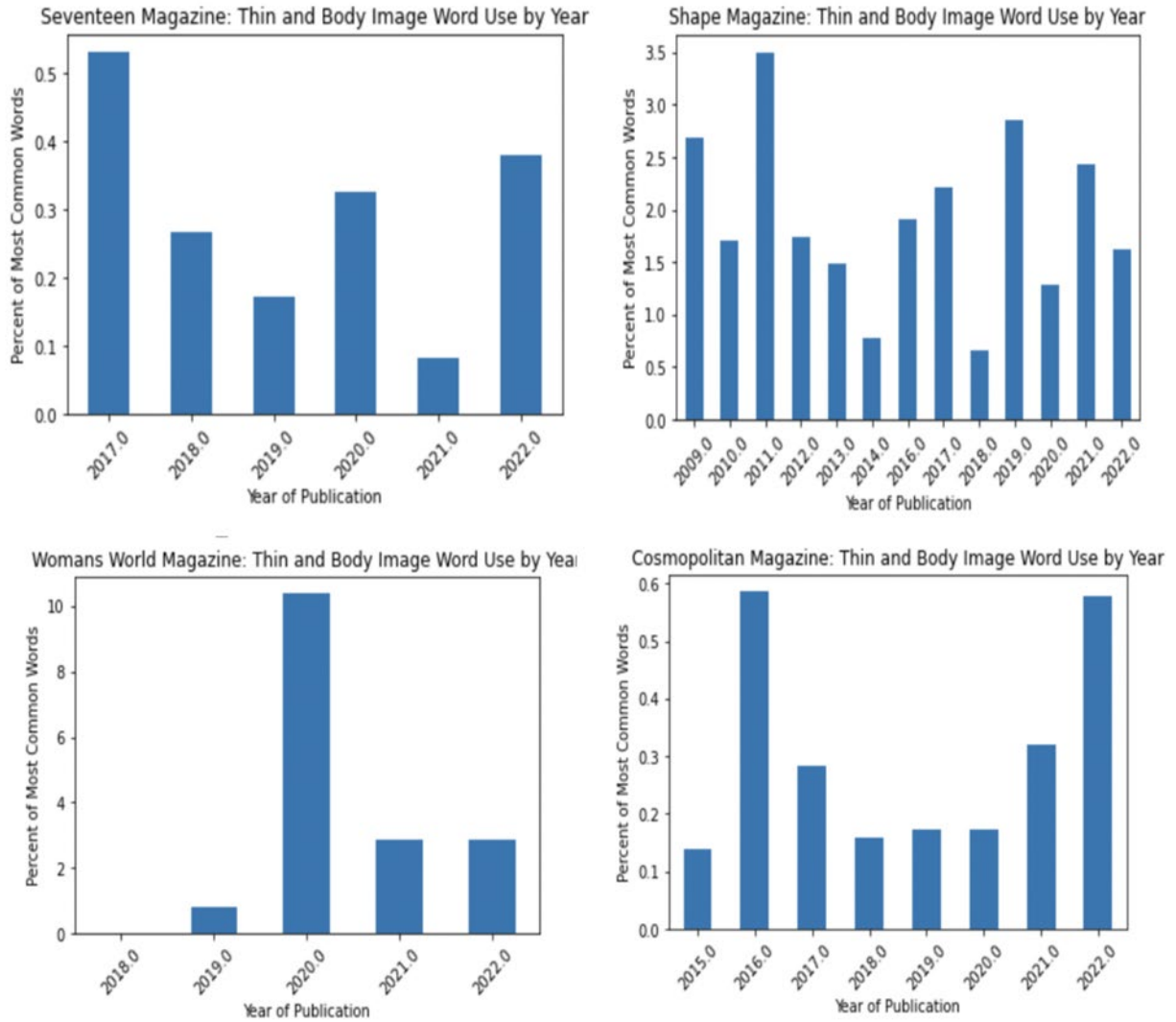
Figure 2. Body-image word frequency by Year

In order to track trends and patterns in the dataset, we decided to track how body-image language was being used over time. This task computed the frequency of body-image words as compared to the most common words across the entire corpus of each women's magazine.

The frequency of body-image language use varies each year. Woman's World has a significant peak usage of body-image language in 2020 of 10%. In comparison to the other years of 2018, 2019, 2021, and 2022, the word frequency is 0% to 3%. For Seventeen, Shape, and Cosmopolitan, there is low variation between years.

Temporal Frequency Tasks by Month
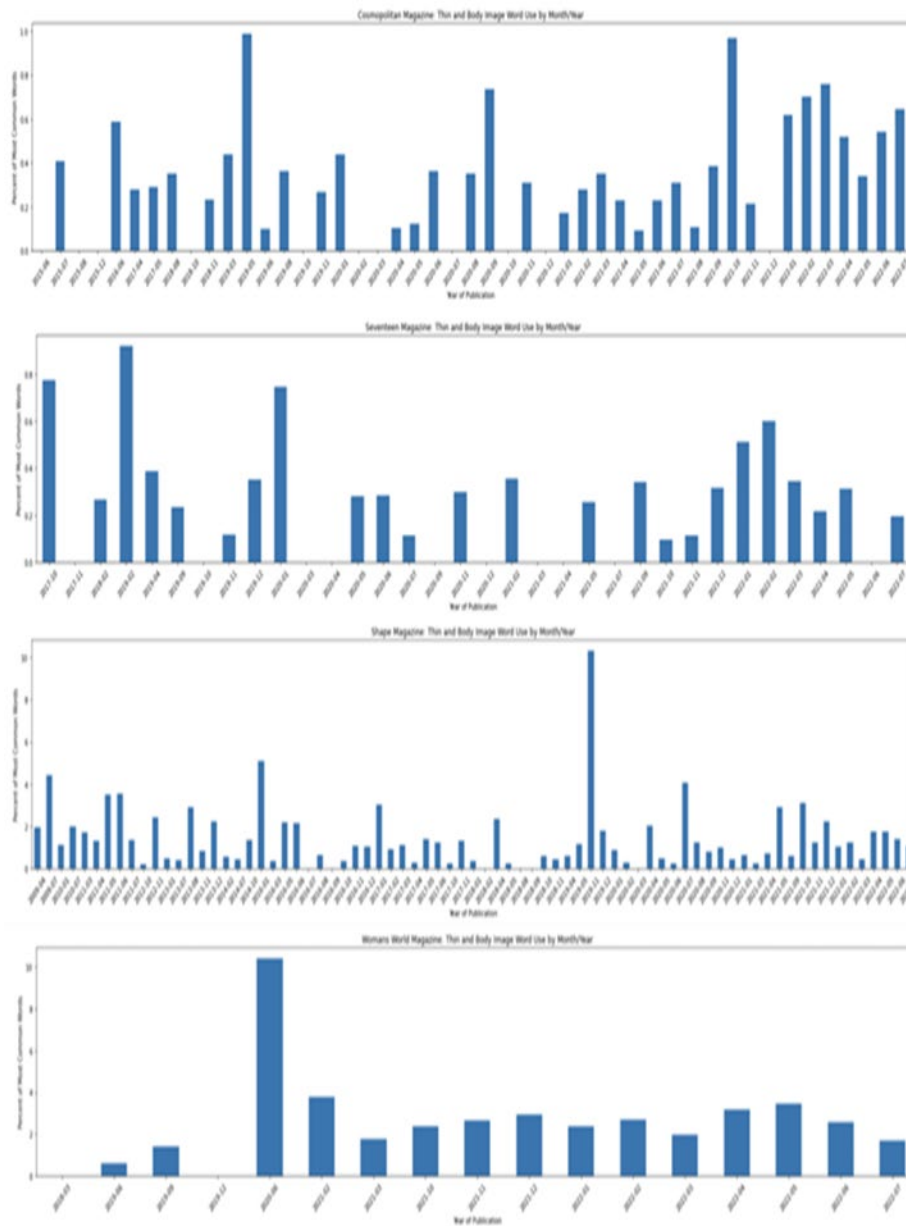
Figure 3. Body image word frequency by month, for each magazine. In order from first appearance to last appearance - Cosmopolitan, Seventeen, Shape, Woman's World.

The frequency of body image language for each magazine varies per month. For Cosmopolitan, there seems to be a higher and steadier frequency of body-related words in early 2022 (0.6%-0.8%). Shape magazine shows denser frequencies of body-image words in 2009-2016 — a range of 1%% to 6% frequencies — before being relatively low, and then a surge appearance again in 2019 — up to 10% in mid-2019. Woman's World shows a high frequency of body words during 2020 (10% in the month of 2020-June), which could be attributed to the COVID-19 pandemic. After 2020, there seems to be a relatively steady frequency of body-image words — a range of 2% to 4% appearance. Seventeen has a low variation.

Comparing The Percentage of Each Word for Each Magazine
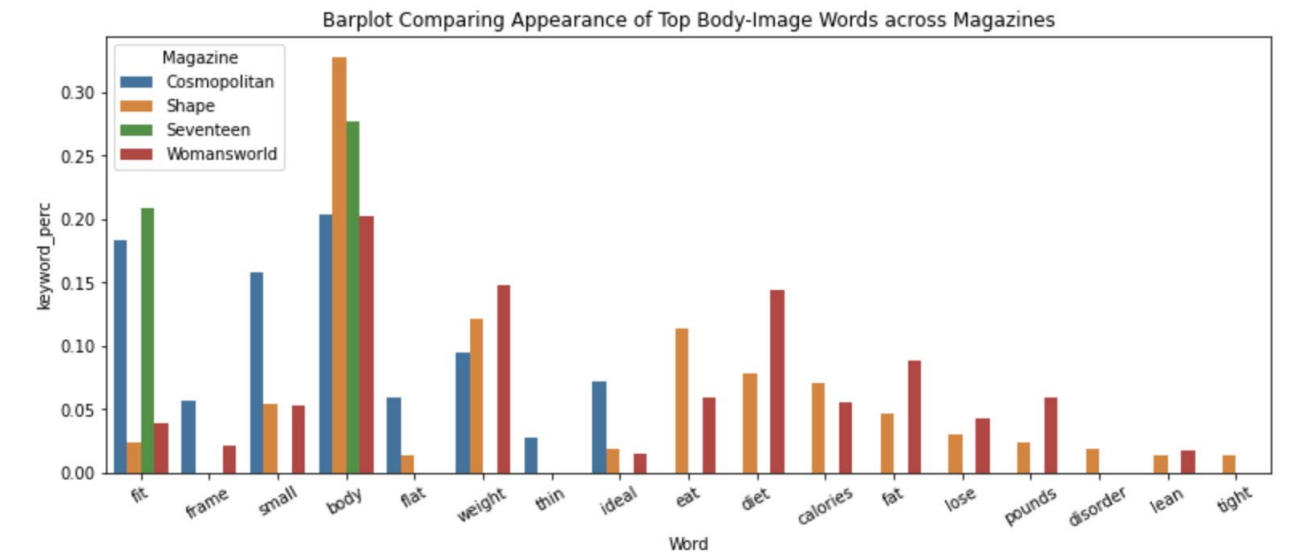
Figure 4. Barplot comparing the frequencies of most used body-image words across magazines

From all four magazines, the word 'body' appears the most out of the keywords. Both Shape and Woman's World have higher frequencies of language related to weight loss or diets, with words such as 'diet', 'eat', 'calories', and 'weight'. This suggests that both Shape and Woman's World have higher usage of language related to weight-loss and diet ideals.

Cosmopolitan has lower frequencies of words such as 'fit', small, 'thin', and 'ideal' as compared to Shape and Woman's World. While there is evidence of Cosmopolitan using body image-related language, lower frequencies may demonstrate that they are not trying to promote the thin ideal or weight loss, as compared to the other magazines. Seventeen does not have a prevalent frequency of body-image words.

Unique Body-Image Words Across Magazines

Seventeen

	Word	Frequency	Keyword	Percentage	Shared
2216	frame	6	True	7.594937	False
2487	weight	6	True	7.594937	False
4024	pounds	3	True	3.797468	False
4805	toned	2	True	2.531646	False
5028	fat	2	True	2.531646	False

Cosmopolitan

	Word	Frequency	Keyword	Percentage	Shared
612	eat	32	True	4.863222	False
1944	pounds	26	True	3.951368	False
2447	slim	16	True	2.431611	False
3467	shrink	9	True	1.367781	False
5949	diet	9	True	1.367781	False

Woman's World

	Word	Frequency	Keyword	Percentage	Shared
231	lose	158	True	8.302680	False
255	fit	145	True	7.619548	False
546	frame	79	True	4.151340	False
798	ideal	57	True	2.995271	False
969	flat	47	True	2.469785	False

Shape

	Word	Frequency	Keyword	Percentage	Shared
498	lose	54	True	6.967742	False
685	pounds	42	True	5.419355	False
710	fit	41	True	5.290323	False
932	ideal	32	True	4.129032	False
1314	flat	23	True	2.967742	False

Figure 5. Table of Top 5 Unique Keywords for Each Magazine

Each table shows keywords most respective and most unique to each magazine. Woman's World and Shape have the highest frequency of words associated with controlling or changing body shape —'flat', 'ideal', 'lose', and 'fit'.

Unique Key-Phrases

Top 5 Unique Body-Image Phrases in Seventeen

	phrase	cnt	is_bodyimage	Magazine	total_cnt	Percentage
16223	your body	30	True	Seventeen	45112	8.196721
13564	the right fit	5	True	Seventeen	45112	1.366120
4234	body	4	True	Seventeen	45112	1.092896
16659	your own body	3	True	Seventeen	45112	0.819672
11809	the body	3	True	Seventeen	45112	0.819672

Top 5 Unique Body-Image Phrases in Cosmopolitan

	phrase	cnt	is_bodyimage	Magazine	total_cnt	Percentage
53416	your body	44	True	Cosmopolitan	193539	4.150943
4629	a flat sheet	25	True	Cosmopolitan	193539	2.358491
45774	the weight	24	True	Cosmopolitan	193539	2.264151
28649	my body	23	True	Cosmopolitan	193539	2.169811
50660	three different weight options	17	True	Cosmopolitan	193539	1.603774

Top 5 Unique Body-Image Phrases in Womans World

	phrase	cnt	is_bodyimage	Magazine	total_cnt	Percentage
28061	the body	223	True	Womansworld	114199	7.384106
37499	your body	217	True	Womansworld	114199	7.185430
36886	weight	150	True	Womansworld	114199	4.966887
9994	calories	104	True	Womansworld	114199	3.443709
37731	your diet	88	True	Womansworld	114199	2.913907

Top 5 Unique Body-Image Phrases in Shape

	phrase	cnt	is_bodyimage	Magazine	total_cnt	Percentage
20665	the body	81	True	Shape	75751	5.756930
14705	my body	48	True	Shape	75751	3.411514
27237	weight	42	True	Shape	75751	2.985075
7613	calories	41	True	Shape	75751	2.914001
11372	her body	30	True	Shape	75751	2.132196

Figure 6. Table of the Top 5 unique body image phrases for each magazine

From all four magazines, the phrase 'your body' or 'the body' has the highest frequency. These tables show similar results to the unique words task. Woman's World has a highest frequency of unique body-image phrases targeted to weight loss or dieting — 'weight', 'calories', 'your diet'. Shape has high use of language related body appearance — 'my body', 'weight', her body'.

Cosmopolitan has a quite standard usage of key phrases, so it is unclear whether they are related to body image or unrelated. Seventeen has little frequencies of body image-related phrases, with only percentages of less than 1.5%.
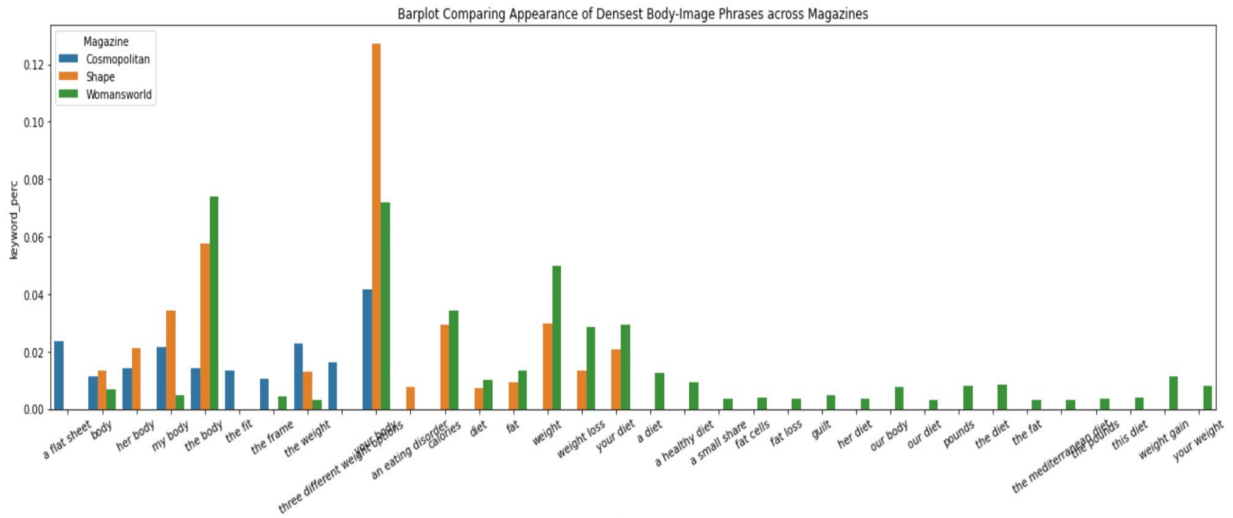
Comparing Density of Body Image Phrases

Figure 7. Barplot comparing appearance of densest body-image phrases across magazines
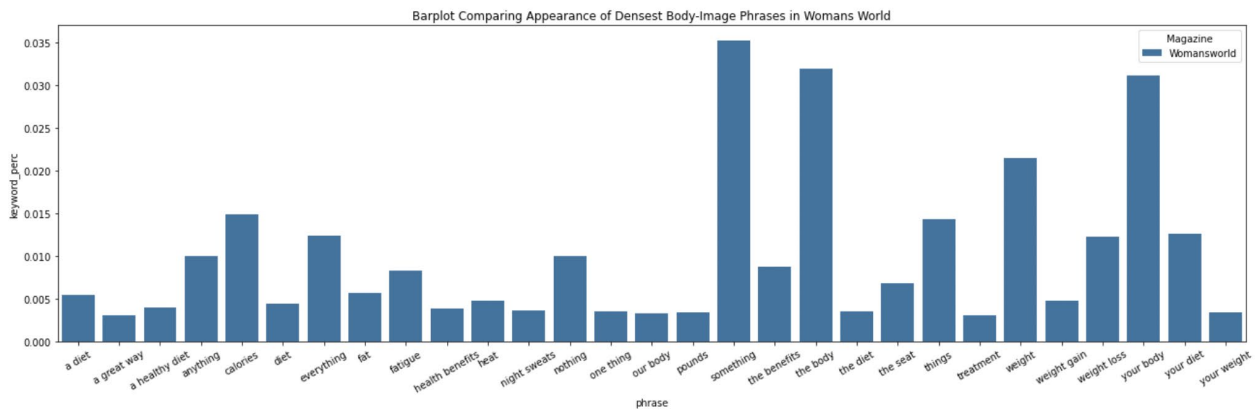

Figure 8. Barplot comparing the appearance of densest body-image phrases in Woman's World

As pictured on the graph, the most frequent appearance of body-image phrases is "your body." This phrase does not specifically have any positive or negative connotations, though it could point out to the fact that Cosmopolitan, Shape, and Woman's World place quite frequent emphasis on body image.

Woman's World continues to have the densest amount of body image-related phrases. There is a frequent amount of phrases related to changing body shape, such as "weight loss", "your diet", "fat loss", etc. While Shape does not have as dense of body-image phrases as Woman's World, they do seem to have frequent usage of phrases with the word "body", as well as "weight loss", "weight options", and "diet."

Discussion

This study provides the relative, quantitative frequencies in which each magazine portrays body image. With a study on the respective frequencies, temporal analysis, and common and unique language used by each magazine, we can get a general, overall measurement of how these magazines discuss body image. Based on the study, Woman's World has high frequencies of body image-related language, with emphasized references to dieting, weight loss, and body shaping. Shape has relatively moderate frequencies of language pertaining to

body image, appearance, and body sculpting. Cosmopolitan and Seventeen have low frequencies of body image language, relative to the other two magazines.

The research has revealed notable variations in the frequency of body-related words among the examined magazines. From the keywords used, the word 'body' is seen to have the highest frequency across all four magazines. Out of all four magazines, Woman's World has the highest frequency and density of language associated with body image. As shown in Fig 1, 1.7% of the total words in Woman's World articles are related to body image. Furthermore, in Fig 7, Woman's World has the highest occurrences of phrases relating to diets, body appearance, and weight loss, in comparison to the other three magazines. The language that Woman's World frequently uses also seems to promote dietary restrictions and methods for weight loss. This suggests that Woman's World frequently targets the message of changing one's body shape, whether to lose weight or fit a certain 'thin' ideal. Woman's World places overwhelmingly high value on glorifying a thin body to their audiences. Shape is marketed as a fitness magazine. Yet, many fitness magazines are notorious for promoting more appearance-related behaviors rather than health-related behaviors (Willis & Knobloch-Westerwick, 2013). Shape's language could suggest the promotion of this appearance-focused behavior.

The study presents the type of body-image words used. Seventeen and Cosmopolitan include synonyms for thin, such as slim and toned. Though it is not certain how these words are used, these words are language associated with encouraging a thinner body image. 'Fit' and 'flat' have become indirect replacements for 'thin', and these words seem to be prevalent in both magazines. On the other hand, Woman's World has a higher percentage of this language with a comparably different audience demographic of middle-aged women. Woman's World may attempt to attract their audience of older women, with articles catered to changing and modifying body shape. In terms of the unique phrases, Woman's World has the highest frequency of language related to dieting and weight loss. Shape uses the highest frequency of phrases targeting body-shaping. This can be attributed to their fitness-targeted market. In fitness and health magazines, messages on weight loss and body sculpting make for over one-fifth of all editorial material. Often, articles on physical activity and health are promoted in connection with attractiveness (Willis & Knobloch-Westerwick, 2013). Messages that perpetuate a certain body ideal, rather than healthy fitness practices, can potentially have a detrimental impact on lowering women's self-esteem, and negatively influencing attitudes regarding body-image.

The study also considers the temporal variations in body image language. In Woman's World, Fig 2 depicts a peak usage of body-image language in 2020. This raises questions about the influence of significant external events in 2020, such as the COVID-19 pandemic, that correlates to the rise of body-image representation. Though Shape, Seventeen, and Cosmopolitan do not have high variations in body image language over time, Shape has a relatively notable, constant variation that could suggest a dynamic engagement with body modifying and fitness-related language. The analysis of unique keywords and phrases highlights the prevalence of body-shaping language in Woman's World and Shape magazines. Both magazines have high frequencies of words associated with controlling or changing body shape, potentially reinforcing unrealistic ideals.

The results of this study could be associated with each magazine's respective audience demographics. Woman's World, catering to a middle-aged audience of 50+ years, exhibits a higher concentration on body image and weight-loss language. This potentially reflects that body image ideals change with age, with older women experiencing greater dissatisfaction with aging and external appearance. Research suggests that more worries about aging prevention in middle-aged people are associated with a deqsire for thinness, body dissatisfaction, and uninhibited eating (Gupta &Schork, 1993; Lewis &Cachelin, 2001) Similarly, a sample of women aged 30-74 in Switzerland reported that 42% had undergone dieting behavior as they aged (Allaz et all.,1998). This suggests that Woman's World could be catering their articles to prevalent body-related behaviors within middle aged women. The quantitative tasks also show that Shape exhibits a high frequency of language pertaining to changing body shape. As a fitness magazine, Shape's audience is those interested in physical fitness, health, and overall well-being. The magazine's content would presumably revolve around topics related to fitness and body health; however, it is still important to consider the underlying contexts behind this language and

whether the magazine is reinforcing toxic body idealizations. On the opposite end, Seventeen, aimed at teenage girls, deviates from body-related articles and hence, has lower frequencies of body-image-related language. Cosmopolitan uses language that could be related to body image, but further research is required to draw conclusive interpretations of their contexts.

Limitations & Future Research

While this study uses frequency tasks, there are limitations to these texts' contextual and semantic analysis. The study does not assess the context and connotations that the keywords are used in. As such, the study might not accurately account for the interpretations, intended messages, and qualitative nuances that could provide a more comprehensive understanding of the magazines' messaging. The study also uses different quantities of corpus data from each magazine, with a possibility of more corpus data on certain dates compared to others. Hence, temporal comparisons may be limited or not provide an accurate representation of the data.

Further computational research could investigate the semantic analysis of these magazines to provide deeper qualitative insight into the intended messages. Additionally, further psychological studies could gather readers' perceptions of the analyzed magazines to explore the impact of the media's body-related language on body image.

Acknowledgments

References

Allaz, A.-F., Bernstein, M., Rouget, P., Archinard, M., & Morabia, A. (1998). Body weight preoccupation in middle-age and ageing women: A general population survey. *International Journal of Eating Disorders*, *23*(3), 287–294. https://doi.org/10.1002/(sici)1098-108x(199804)23:3%3C287::aid-eat6%3E3.0.co;2-f

Aparicio-Martinez, P., Perea-Moreno, A.-J., Martinez-Jimenez, M. P., Redel-Macías, M. D., Pagliari, C., & Vaquero-Abellan, M. (2019). Social Media, Thin-Ideal, Body Dissatisfaction and Disordered Eating Attitudes: An Exploratory Analysis. *International Journal of Environmental Research and Public Health*, *16*(21), 4177. https://doi.org/10.3390/ijerph16214177

Conlin, L., & Bissell, K. (2014). Beauty Ideals in the Checkout Aisle: Health-Related Messages in Women's Fashion and Fitness Magazines. *Journal of Magazine Media*, *15*(2). https://doi.org/10.1353/jmm.2014.0004

Guptaxy, M. A., & Schork, N. J. (1993). Aging-related concerns and body image: Possible future implications for eating disorders. *International Journal of Eating Disorders*, *14*(4), 481–486. https://doi.org/10.1002/1098-108x(199312)14:4%3C481::aid-eat2260140411%3E3.0.co;2-g

Lewis, D. M., & Cachelin, F. M. (2001). Body Image, Body Dissatisfaction, and Eating Attitudes in Midlife and Elderly Women. *Eating Disorders*, *9*(1), 29–39. https://doi.org/10.1080/106402601300187713

Swiatkowski, P. (2016). Magazine influence on body dissatisfaction: Fashion vs. health? *Cogent Social Sciences*, *2*(1). https://doi.org/10.1080/23311886.2016.1250702

Sypeck, M. F., Gray, J. J., & Ahrens, A. H. (2004). No longer just a pretty face: Fashion magazines' depictions of ideal female beauty from 1959 to 1999. *International Journal of Eating Disorders*, *36*(3), 342–347 https://doi.org/10.1002/eat.20039

Willis, L. E., & Knobloch-Westerwick, S. (2013). Weighing Women Down: Messages on Weight Loss and Body Shaping in Editorial Content in Popular Women's Health and Fitness Magazines. *Health Communication*, *29*(4), 323–331. https://doi.org/10.1080/10410236.2012.755602