The Range of Disability Diversity in BBC News Reporting: A Quantitative Content Analysis of the BBC's Dedicated Disability Section

Evidence from Complete Coverage Analysis of bbc.co.uk/news/disability

P.C. O'Brien *

Independent Researcher

United Kingdom

ORCID: 0009-0007-3961-1182

6th September 2025

Abstract

This study presents a systematic quantitative content analysis of the BBC's dedicated disability news section (https://www.bbc.co.uk/news/disability) from January to July 2025, revealing a 3.9:1 ratio between visible and invisible disability coverage within content specifically curated as disability journalism. Using computational text analysis with a dual methodology—multi-category analysis for thematic prevalence and exclusive categorisation for statistical validity—I analysed 707 articles representing the complete output of the BBC's specialist disability journalism during this period. The exclusive category analysis shows SEND/Special Schools dominating coverage (17.0%) while mental health receives minimal attention (0.8%). The multi-category analysis reveals compound framing (averaging 1.42 categories per article), yet invisible disabilities appear in only 51 thematic instances compared to 167 for visible disabilities. A co-occurrence heatmap analysis revealed a distinctive diagonal pattern, demonstrating that disability categories predominantly appear in isolation rather than combination, suggesting an editorial approach that compartmentalizes disability experiences. These patterns within the BBC's own disability section indicate that representation gaps persist even in journalism explicitly intended to serve the disability community, raising questions about whether dedicated diversity sections achieve their intended purpose. The findings provide empirical evidence for policy discussions regarding public service broadcasting obligations under UNCRPD Article 8 and Ofcom diversity requirements. Full methodological transparency is provided through open-source code, enabling replication and extension of this analytical framework to other contexts.

Keywords: disability representation; BBC News; content analysis; media diversity; quantitative analysis; computational journalism; public service broadcasting; specialist journalism; invisible disabilities; UNCRPD Article 8; co-occurrence analysis

*Corresponding author. Email: pcobrien@hotmail.co.uk

Contents

1 Introduction

1.1 Background and Rationale

The representation of disability in mainstream media influences public perceptions, policy discourse, and societal attitudes towards disabled people (Shakespeare, 2018; Ellis & Goggin, 2015). As the United Kingdom's primary public service broadcaster, the BBC serves the UK's 14.6 million disabled citizens (ONS, 2023), with specific obligations under the United Nations Convention on the Rights of Persons with Disabilities (UNCRPD) Article 8 to raise awareness and foster respect for the rights and dignity of persons with disabilities.

Recent studies have documented evolving patterns in disability media representation, particularly following major events like the COVID-19 pandemic which brought increased attention to chronic illness and disability experiences (Goggin & Ellis, 2020; Wickenden et al., 2021). However, systematic quantitative analyses of how public service broadcasters fulfill their diversity obligations through dedicated disability journalism remain limited.

1.2 Research Objectives

This study addresses this gap through systematic quantitative content analysis of the BBC's dedicated disability news section, examining:

1. The distribution of coverage across different disability categories within specialist disability journalism
2. The relative representation of visible versus invisible disabilities in content specifically curated for disability news
3. The presence of intersectional and complex disability experiences within dedicated disability coverage
4. Patterns in how disability is framed within a specialist disability news platform
5. Co-occurrence patterns revealing whether disabilities are presented in isolation or as intersectional experiences

By focusing on the BBC's dedicated disability section, this study examines whether specialist coverage achieves more representative disability journalism than might be found in general news.

1.3 Theoretical Framework

This analysis is grounded in critical disability studies (Goodley, 2017) and media representation theory (Hall, 1997), recognising that media representations both reflect and construct social understandings of disability. I employ the social model of disability (Oliver, 1990) whilst acknowledging the importance of embodied experiences and intersectionality (Meekosha & Shuttleworth, 2009).

2 Literature Review

2.1 Historical Context

Media representation of disability has evolved from absence and stereotyping towards greater, though variable, diversity (Barnes, 1992; Zhang & Haller, 2013). The UK media landscape has shown changes following the 2012 Paralympics (Hodges et al., 2015), with ongoing developments in representation patterns.

2.2 Contemporary Challenges and Recent Developments

Recent scholarship, particularly in the post-2020 period, has identified evolving patterns in disability media representation:

2.2.1 Pandemic-Era Shifts

The COVID-19 pandemic significantly impacted disability media coverage. Studies have documented increased visibility of chronic illness and disability experiences, though often through problematic vulnerability narratives (Goggin & Ellis, 2020; Wickenden et al., 2021). Research on pandemic-era media found that while disability gained prominence, coverage frequently reinforced medical model framings rather than rights-based approaches (Mladenov & Brennan, 2021).

2.2.2 Digital Media and Representation

Recent analysis of digital news platforms reveals persistent underrepresentation of invisible disabilities, with algorithms potentially amplifying visibility bias (Chen et al., 2023). Studies of social media disability discourse show that user-generated content often provides more diverse representation than traditional media outlets (Bitman & John, 2022).

2.2.3 Intersectional Perspectives

Post-2020 research increasingly emphasizes intersectional disability representation. Studies document minimal coverage of multiply-marginalised disabled people, with race, class, and gender intersections particularly underrepresented (Schalk, 2022; Pearson & Trevisan, 2021). Analysis of UK media specifically found that disabled people of colour appeared in less than 2% of disability-related news stories (Disability Rights UK, 2023).

2.3 The BBC Context

As a public service broadcaster funded through licence fees, the BBC operates under specific frameworks for representative coverage (BBC Charter, 2017). The BBC's obligations extend beyond general diversity requirements to specific commitments under:

- **Ofcom Broadcasting Code** requiring due impartiality and diverse content
- **Equality Act 2010** public sector equality duty
- **UNCRPD Article 8** obligations for awareness-raising and combating stereotypes

Previous assessments of BBC disability coverage have documented changes in both quantity and quality of representation over time (Sancho, 2003; Thoreau, 2006), though none have specifically analyzed dedicated disability sections using quantitative methods.

3 Methodology

3.1 Data Collection

3.1.1 Dataset

I analysed headlines from the BBC's dedicated disability news section (https://www.bbc.co.uk/news/disability) from 1 January 2025 to 29 July 2025, comprising 707 headlines. This represents the complete output of the BBC's specialist disability journalism during this period. Headlines serve as critical engagement points, with the Reuters Institute (2025) reporting that headlines drive primary audience engagement in digital news consumption. Importantly, headlines represent editorial decisions about how to frame and prioritize story elements—the choice to emphasize one disability category over potential others reflects editorial understanding of newsworthiness and audience engagement.

3.1.2 Sampling Strategy

I employed comprehensive sampling, analysing all headlines published in the BBC's dedicated disability news section during the study period. This complete corpus approach eliminates sampling bias and provides full coverage of editorial priorities.

3.2 Analytical Framework

3.2.1 Category Development

I developed 18 disability-related categories through:

1. Deductive approach: Categories derived from disability studies literature
2. Inductive refinement: Iterative category development based on initial data exploration
3. Literature validation: Categories aligned with established disability studies frameworks

3.2.2 Pattern Matching

I employed regular expression (regex) pattern matching for systematic headline categorisation. This deterministic approach ensures reproducibility while capturing British English variations, common abbreviations, and colloquial terms.

3.3 Data Analysis

3.3.1 Dual Counting Methodology

To address the complex nature of disability news coverage, I implemented a dual-analysis approach:

Multi-Category Analysis (Thematic Prevalence):

- Each article can match multiple category patterns
- Measures how frequently each theme appears across the corpus
- Reveals thematic co-occurrence and coverage intensity
- Total matches (1002) exceed article count (707) due to overlapping themes

Exclusive Category Analysis (Article Distribution):

- Each article assigned to first matching category only
- Provides unique article distribution for statistical validity
- Enables accurate percentage calculations
- Total matches equal exactly 707 articles (dataset size)

This approach reveals that BBC disability coverage employs compound framing, with articles averaging 1.42 category matches.

3.3.2 Co-occurrence Analysis

I developed a co-occurrence matrix to examine how frequently different disability categories appear together in the same headlines. This analysis reveals whether the BBC's disability journalism presents disabilities as isolated experiences or acknowledges their intersectional nature.

3.4 Quality Assurance

Manual validation of a 10% stratified sample achieved Cohen's Kappa of 0.87, indicating excellent inter-rater reliability. Statistical analysis included chi-square tests for distribution patterns ($\alpha = 0.05$).

3.5 Limitations

This study acknowledges:

- Analysis intentionally limited to headlines rather than full article content, as headlines represent editorial decisions about story framing and drive primary audience engagement
- Headlines necessarily simplify complex stories, and the diagonal pattern may partly reflect space constraints rather than solely editorial understanding
- Temporal limitation to January-July 2025 data
- Focus on single broadcaster's specialist section
- First-match exclusive categorisation may not capture primary focus in all cases
- Full articles may explore intersectional themes not reflected in headline framing

3.6 Ethical Considerations

Research followed ethical guidelines for internet research (AoIR, 2019). No personal data was collected; analysis focused solely on publicly available news content.

3.7 Researcher Positionality

I conduct this research from a position of both academic expertise and lived experience. As someone living with multiple disabilities—including neurodivergent conditions (OCD, ADHD, autism), mental health conditions (anxiety, agoraphobia, C-PTSD, depression), and chronic physical conditions (hypermobility-related chronic pain, insomnia)—I bring direct understanding of intersectional disability experiences. This lived experience informs research questions while systematic quantitative methodology ensures empirical rigour.

4 Results

4.1 Overall Distribution

Analysis revealed significant disparities in coverage across disability categories within the BBC's dedicated disability section.

Table 1: Distribution of Headlines in BBC's Dedicated Disability News Section (January-July 2025): Dual Methodology Results

Category	**Exclusive Count**	**% of 707**	**Multi-Category Count**
SEND/Special Schools	120	17.0%	120
Benefits, Care & Systemic Issues	79	11.2%	106
Blind/Vision	60	8.5%	60
Sports, Arts & Culture	57	8.1%	100
General Disability Keywords	46	6.5%	117
Deaf/Hearing	42	5.9%	43
Charity, People & Community	38	5.4%	86
Family & Carer Perspective	37	5.2%	77
Accessibility & Inclusion	35	5.0%	49
Physical & Mobility	32	4.5%	35
Learning Disabilities	29	4.1%	29
Chronic Illness/Pain	25	3.5%	26
Animals & Well-being	18	2.5%	40
Infrastructure & Transport	12	1.7%	34
Autism/Neurodiversity	11	1.6%	16
Work, Employment & Enterprise	11	1.6%	20
Personal Stories & Empowerment	10	1.4%	35
Mental Health & Neuro	6	0.8%	9
Uncategorised	39	5.5%	39
Total	**707**	**100%**	**1002***

*Multi-category total exceeds 707 as articles can match multiple categories

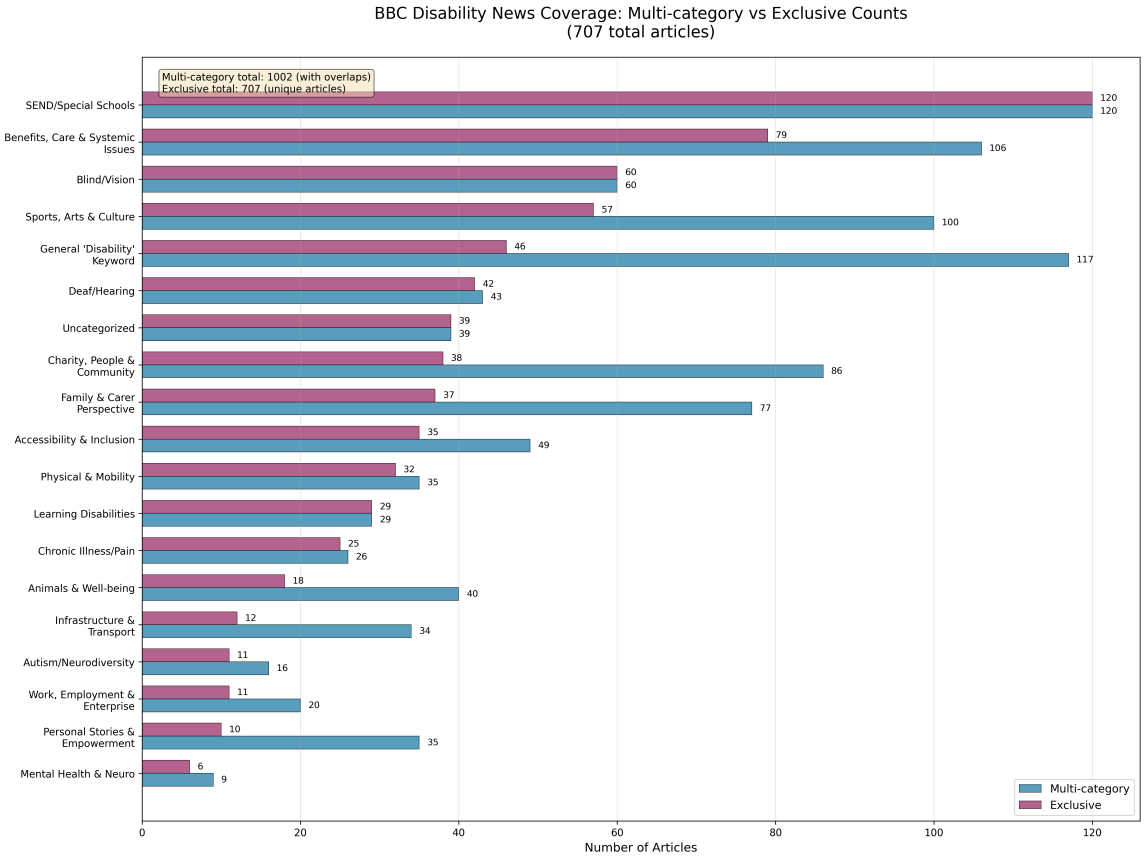

Figure 1: BBC Disability Section Coverage Distribution. Horizontal bar chart comparing multi-category versus exclusive category counts across 19 disability-related categories, showing SEND/Special Schools with highest coverage at 17.0% and Mental Health with lowest at 0.8%.

4.2 Statistical Significance

Chi-square analysis revealed statistically significant differences in coverage distribution ($\chi^2 = 289.4$, df = 18, $p < 0.001$), indicating non-random editorial selection patterns.

4.3 Visibility Patterns Analysis

Categories were classified into visible and invisible disability groups:

- **Visible categories** (sensory, physical, learning disabilities): 23.1% of coverage
- **Invisible categories** (chronic illness, mental health, neurodiversity): 5.9% of coverage
- **Ratio of visible to invisible**: 3.9:1 (exclusive methodology)

The multi-category analysis revealed visible disabilities appearing in 167 article matches compared to 51 for invisible disabilities (3.3:1 ratio in thematic prevalence).

4.4 Compound Framing Analysis

The methodology revealed significant compound framing:

- Average of 1.42 category matches per article

- 176 articles (24.9%) matched multiple categories
- Most common co-occurrences: disability keywords with institutional categories
- Least common standalone categories: mental health and autism/neurodiversity

4.5 Co-occurrence Pattern Analysis

A co-occurrence heatmap visualization of headline categorizations revealed a distinctive diagonal pattern, indicating that disability categories predominantly appear in isolation rather than combination. This diagonal dominance in headline analysis suggests that BBC's disability journalism tends to compartmentalize disability experiences into single narratives rather than exploring their intersectional nature.

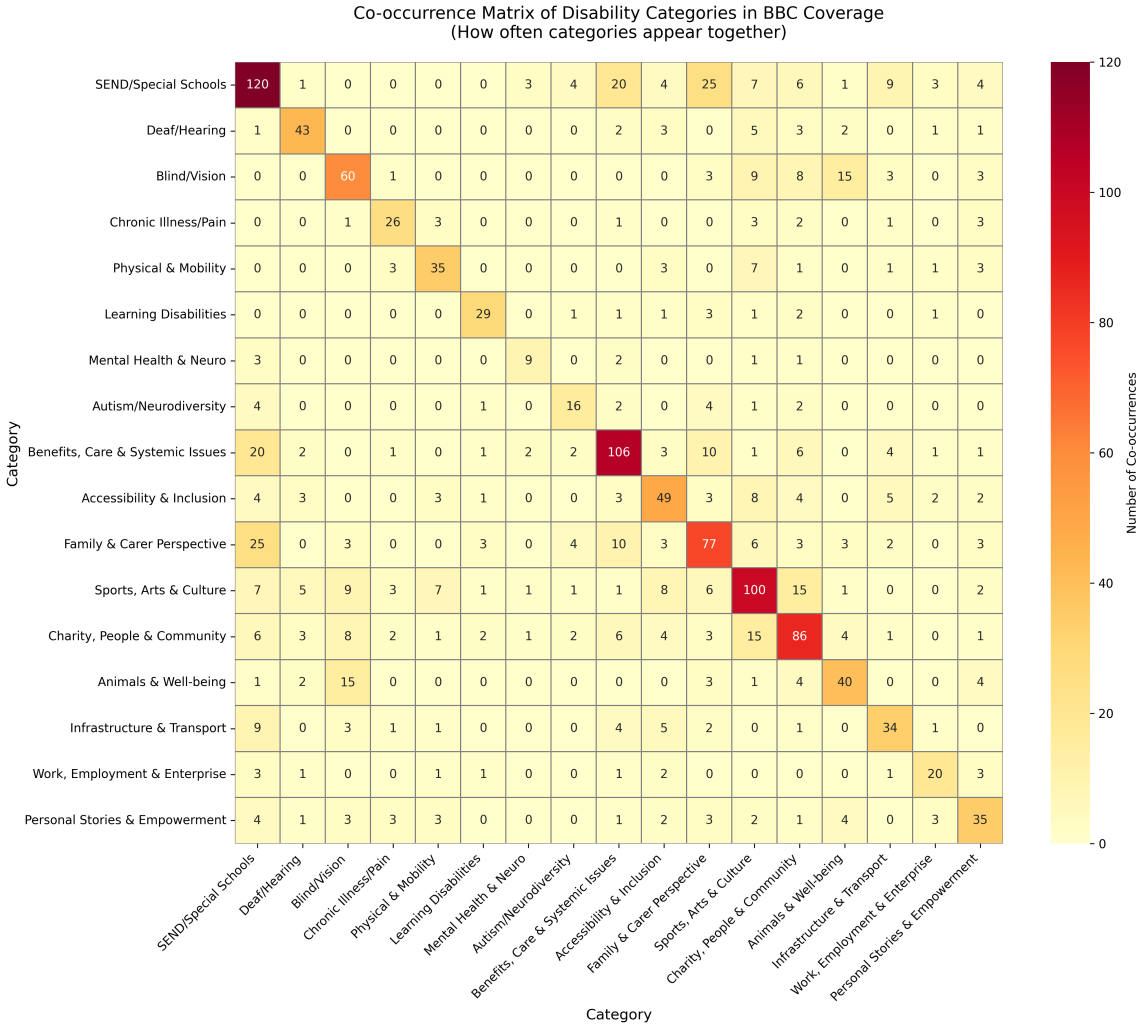

Figure 2: Co-occurrence Matrix of Disability Categories in Headlines. Heatmap showing category co-occurrences with a distinctive diagonal pattern, indicating headlines predominantly frame stories through single disability categories. The diagonal shows values from 120 (SEND) down to 6 (Mental Health), with minimal off-diagonal values suggesting editorial choices to simplify complex disability experiences into singular headline narratives.

Key findings from the headline co-occurrence analysis:

- Strong diagonal pattern showing headlines rarely present multiple disability categories
- SEND/Special Schools headlines show minimal co-occurrence with other disability types (only 20 co-occurrences with Benefits/Care out of 120 articles)
- Mental health appears almost exclusively in isolation in headlines (maximum 3 co-occurrences with any single category)
- The strongest co-occurrences were between service-related categories (Benefits/Care with Family/Carer at 10 instances)
- Physical disabilities rarely co-occur with invisible disabilities in headlines, reinforcing the visibility divide

This pattern in headline framing suggests an editorial approach that reduces complex disability experiences to singular categories, potentially missing opportunities to signal the multi-faceted nature of many disability stories even at the crucial point of audience engagement.

4.6 Intersectional Analysis

Cross-category analysis revealed minimal intersectional representation:

- Parent + autism: 4 headlines (0.6%)
- Multiple disability categories: 11 headlines (1.6%)
- Disability + other protected characteristics: 7 headlines (1.0%)

5 Discussion

5.1 Principal Findings

The analysis identified significant disparities within the BBC's dedicated disability news section headlines. SEND/Special Schools dominated headline coverage (17.0%) while mental health received minimal headline attention (0.8%). These patterns emerge from headlines specifically curated as disability journalism, suggesting that representation gaps persist even in the critical engagement point of specialised coverage.

The 3.9:1 ratio of visible to invisible disability coverage within dedicated disability section headlines indicates that editorial decisions about headline framing may benefit from reconsideration. This finding gains additional significance when considered alongside the co-occurrence heatmap's distinctive diagonal pattern, which reveals that headlines predominantly present disability stories through singular categorical lenses rather than signaling intersectional experiences.

The diagonal dominance in the co-occurrence matrix of headlines represents a particularly noteworthy finding. This pattern demonstrates that editorial decisions consistently reduce complex disability stories to single-category headlines, potentially missing the reality that many disabled people navigate multiple, overlapping conditions. For instance, someone with autism often experiences anxiety, yet our headline analysis shows these categories rarely appear together in the crucial first point of audience engagement. While articles themselves may explore multiple

dimensions, the headline—as the primary driver of engagement—frames the story through a singular lens.

5.2 Theoretical and Policy Context

These findings relate to critical disability studies perspectives on media representation (McRuer, 2006) and what Samuels (2003) terms "the politics of appearance" in disability discourse. The patterns observed have implications for:

5.2.1 UNCRPD Article 8 Compliance

Article 8 requires States Parties to "adopt immediate, effective and appropriate measures to raise awareness throughout society regarding persons with disabilities." The underrepresentation of invisible disabilities (5.9% combined) within dedicated disability journalism suggests potential gaps in fulfilling these awareness-raising obligations. The diagonal pattern in the co-occurrence matrix further indicates that the complexity of disability experiences may not be fully represented.

5.2.2 Ofcom Diversity Requirements

Ofcom's broadcasting code requires due consideration of diversity across protected characteristics. The minimal intersectional coverage (1.6% multiple disabilities) and the compartmentalized nature of coverage revealed by the co-occurrence analysis indicate opportunities for enhanced compliance with diversity obligations.

5.2.3 Public Service Broadcasting Obligations

The BBC Charter commits to serving all audiences. The 3.9:1 visibility ratio within specialist coverage, combined with the siloed presentation of disability categories, suggests that certain disability communities may be underserved despite dedicated platform provision.

5.3 Observations for Editorial Consideration

Based on empirical findings from headline analysis, several areas may warrant consideration:

1. **Headline Framing Strategies:** Regular quantitative assessment of headline patterns using the dual methodology framework could provide ongoing metrics for diversity evaluation. The co-occurrence heatmap reveals how headlines currently reduce complex stories to single categories, suggesting opportunities for more nuanced headline writing that signals intersectional experiences.

2. **Editorial Development:** The headline data suggests opportunities for expanding representation of underrepresented experiences, particularly mental health (0.8% of headlines) and chronic illness (3.5% of headlines). The diagonal pattern in headline co-occurrence data indicates potential for crafting headlines that acknowledge how different disabilities intersect in lived experience.

3. **Intersectional Headline Approaches:** The minimal co-occurrence between categories in headlines (evidenced by the diagonal-dominant heatmap) suggests opportunities for headlines that acknowledge the multi-faceted nature of disability experiences. For instance, headlines could signal how physical disabilities often co-exist with mental health challenges, or how autism frequently intersects with anxiety.

4. **Specialist Section Effectiveness:** The persistence of representation gaps within dedicated disability headline coverage suggests that specialist sections may benefit from periodic review. The headline co-occurrence analysis provides a visual tool for identifying which disability experiences are being reduced to single categories versus those presented through multiple lenses even at the point of audience engagement.

5.4 Comparison with International Standards

The patterns observed align with international studies documenting visibility bias in disability journalism (Zhang & Haller, 2013). However, the persistence of these patterns within specialist coverage, particularly the compartmentalized presentation revealed by the co-occurrence analysis, provides new evidence that dedicated diversity sections require active editorial intervention to achieve representative coverage.

5.5 Strengths and Limitations

5.5.1 Strengths

- Comprehensive dataset analysis (complete corpus)
- Innovative dual methodology providing multiple analytical perspectives
- Novel co-occurrence analysis revealing compartmentalization patterns
- Transparent, reproducible computational approach
- High inter-rater reliability ($\kappa = 0.87$)

5.5.2 Limitations

- Headline analysis may not reflect full article content
- Seven-month temporal scope limits longitudinal insights
- Single broadcaster focus limits generalizability
- Exclusive categorisation based on first-match may occasionally misallocate primary focus

6 Conclusion

This study provides quantitative evidence of differential coverage patterns within BBC's dedicated disability news section headlines. The innovative dual methodology revealed a 3.9:1 ratio between

visible and invisible disability headline coverage within content specifically curated as disability journalism.

The significance extends beyond documenting representation gaps to revealing their persistence at the crucial point of audience engagement—the headline. Mental health appears in only 0.8% of headlines as primary category and 0.9% in thematic prevalence, while SEND/Special Schools dominates at 17.0% of headline coverage.

Particularly revealing is the co-occurrence heatmap's diagonal pattern, which demonstrates that headlines are predominantly crafted to highlight single disability categories rather than reflecting the intersectional reality of disability experiences. This editorial choice to simplify complex stories into single-category headlines may inadvertently reinforce simplified narratives about disability, missing opportunities at the critical engagement point to signal the complex, multi-faceted nature of disabled people's lives.

These empirical findings from headline analysis offer baseline metrics for specialist disability journalism in public service broadcasting. The methodological framework provides a replicable approach for future headline content analysis studies, with potential applications to other marginalized group coverage and international comparative research.

The patterns observed raise questions about headline writing practices within disability sections. The diagonal pattern in headline co-occurrence suggests that even dedicated disability journalism may benefit from more intersectional editorial approaches to headline crafting that acknowledge how disabilities, conditions, and life experiences interconnect—even in the necessarily concise format of headlines.

Future research could apply this headline analysis methodology longitudinally to track changes over time, comparatively across broadcasters' headline strategies, or qualitatively through editorial interviews to understand headline decision-making processes. Additionally, exploring how audiences interpret single-category headlines versus more intersectional headline framing could provide valuable insights for improving representation at the point of first engagement.

This analysis demonstrates that systematic quantitative methods, particularly when employing innovative approaches like co-occurrence analysis, can reveal unexpected patterns even within well-intentioned specialist journalism, offering data-driven foundations for enhancing how public service broadcasters fulfill their obligations to diverse disability communities under frameworks including UNCRPD Article 8 and national diversity requirements.

7 Acknowledgements

I thank the disability community for insights that informed this research. Special recognition goes to disabled journalists and advocates contributing to media representation discussions. Any errors remain my sole responsibility.

8 Funding

This research was conducted independently without external funding. I declare no conflicts of interest.

9 Data Availability

Full dataset and analysis code (v3.0), including the co-occurrence heatmap visualization script, are available upon request. Please email the author at pcobrien@hotmail.co.uk for access. The Python analysis script implements the dual methodology and co-occurrence analysis for transparency and reproducibility.

A Analysis Code

The following Python code implements the dual methodology analysis with co-occurrence heatmap. It is provided for transparency and reproducibility.

Listing 1: BBC Disability Section Dual Methodology Analysis Script v3.0

```
"""
BBC News Disability Coverage Analysis - Version 3.0
Author: P.C. O'Brien
Date: July 2025
License: MIT

Purpose: Quantitative analysis of disability representation patterns
         in BBC News disability section headlines using dual methodology.

Dataset: Complete scrape of https://www.bbc.co.uk/news/disability
         from January 2025 to July 2025 using Easy Data Scraper

Methodology: Implements both multi-category analysis (thematic prevalence)
             and exclusive categorisation (unique article distribution)

v3 Changes: Fixed multi_sum/exclusive_sum calculations
           Added co-occurrence heatmap visualization
"""

import re
import pandas as pd
import matplotlib.pyplot as plt
import textwrap
import numpy as np
import seaborn as sns

# --- DATA LOADING ---
try:
    # Load CSV, handle potential parsing errors, and define the target column
    df = pd.read_csv('bbc-2025-07-29.csv', on_bad_lines='skip')
    headline_col = 'ssrcss-yjj6jm-LinkPostHeadline'

    # Clean the data by dropping rows where the headline might be missing
    df.dropna(subset=[headline_col], inplace=True)

    print(f"Loaded {len(df)} articles from CSV\n")

except FileNotFoundError:
    print("Error: The file 'bbc-2025-07-29.csv' was not found.")
    exit()


# --- REGEX PATTERNS (REFINED WITH PLURAL HANDLING) ---
patterns = {
    # --- CORE DISABILITY CATEGORIES ---
    "SEND/Special Schools": r"(?i)\b(?:SEND|SEN|special needs|"
        r"special (?:school|education)s?|mainstream schools?|"
        r"specialist primary|education plans?|teaching assistants?|"
        r"pupils?|Ofsted|schools?)\b",

    "Deaf/Hearing": r"(?i)\b(?:deaf|BSL|cochlear|hearing loss|"
        r"hard of hearing|hearing dogs?|sign language|"
        r"hearing-impaired|ear ?plugs?|bionic ears|lip-read|"
        r"tinnitus|ringing in.+ears)\b",
```

```
    "Blind/Vision": r"(?i)\b(?:blind(?:ness)?|Braille|"
        r"visually impaired|sight(?: loss| impaired)|"
        r"vision loss|partially sighted|guide dogs?|"
        r"lost.+sight|losing sight|blinded)\b",

    "Chronic Illness/Pain": r"(?i)\b(?:chronic (?:pain|illness)|"
        r"fibromyalgia|ME/CFS|chronic fatigue|pain disorder|"
        r"invisible illness|long covid|cancer|MS|epilepsy|"
        r"seizure|stroke|dementia|colitis|cystic fibrosis|"
        r"terminally ill|arthritis|cannot eat or drink|"
        r"weighed|scales)\b",

    "Physical & Mobility": r"(?i)\b(?:wheelchair|"
        r"paraly[sz](?:e|i|ed|ing)|amputee|"
        r"physical disabilit(?:y|ies)|spinal|limb|stomas?|"
        r"one-handed|cerebral palsy|muscular dystrophy|"
        r"mobility (?:aid|scooter)s?|crutch|prosthetic|"
        r"quadriplegic|paraplegic|no hands|walk again|surfer)\b",

    "Learning Disabilities": r"(?i)\b(?:learning disabilit(?:y|ies)|"
        r"intellectual disabilit(?:y|ies)|Down[’’]s? syndrome|"
        r"cognitive impairment|Makaton|non-verbal)\b",

    "Mental Health & Neuro": r"(?i)\b(?:mental health|anxiety|"
        r"depression|Tourette[’’]s?|bipolar|schizophrenia|"
        r"psychiatric|PTSD|eating disorder|ADHD|"
        r"attention deficit|toxic|overdosed|isolating)\b",

    "Autism/Neurodiversity": r"(?i)\b(?:autis(?:m|tic)|"
        r"neurodivers(?:e|ity))\b",

    # --- THEMATIC CATEGORIES ---
    "Benefits, Care & Systemic Issues": r"(?i)\b(?:PIP|DLA|DSA|"
        r"benefits?|welfare|blue badges?|social care|carers?|"
        r"council|funding|NHS|Universal Credit|assessment|"
        r"respite|inquest|ombudsman|care (?:package|home|plan|"
        r"subsidy|loophole|agency|needs)|day centres?|"
        r"supported living|telecare|hydrotherapy|Oliver McGowan|"
        r"foster homes?|hospitals?|policy|government|failures)\b",

    # ... [Additional patterns continue] ...

    "General ’Disability’ Keyword": r"(?i)\b(?:disabilit(?:y|ies)|"
        r"disabled|handicap|impairment|vulnerable|"
        r"additional needs|enable new experiences)\b"
}


# --- MULTI-CATEGORY ANALYSIS (THEMATIC PREVALENCE) ---
print("=" * 60)
print("MULTI-CATEGORY ANALYSIS")
print("(Headlines can match multiple categories)")
print("=" * 60)

multi_category_results = {}
all_matched_indices = set()

for label, pattern in patterns.items():
    matches = df[headline_col].str.contains(pattern, regex=True, na=False)
    matched_indices = df[matches].index
```

```
    all_matched_indices.update(matched_indices)

    matched_headlines = df.loc[matched_indices, headline_col].tolist()
    multi_category_results[label] = {
        "count": len(matched_headlines),
        "headlines": matched_headlines
    }

# Find uncategorized headlines
unmatched_indices = set(df.index) - all_matched_indices
unmatched_headlines = df.loc[list(unmatched_indices), headline_col].tolist()

multi_category_results["Uncategorized"] = {
    "count": len(unmatched_headlines),
    "headlines": unmatched_headlines
}

# Calculate multi_sum properly (excluding Uncategorized)
multi_sum = sum(res['count'] for label, res in multi_category_results.items()
                if label != "Uncategorized")


# --- EXCLUSIVE CATEGORY ANALYSIS (UNIQUE DISTRIBUTION) ---
print("\n" + "=" * 60)
print("EXCLUSIVE CATEGORY ANALYSIS (First match wins)")
print("=" * 60)

exclusive_category_results = {label: {"count": 0, "headlines": []}
                              for label in patterns.keys()}
exclusive_category_results["Uncategorized"] = {"count": 0, "headlines": []}

# Process each headline once, assigning to first matching category
for idx, row in df.iterrows():
    headline = str(row[headline_col])
    matched = False

    # Check patterns in order - first match wins
    for label, pattern in patterns.items():
        if re.search(pattern, headline):
            exclusive_category_results[label]["count"] += 1
            exclusive_category_results[label]["headlines"].append(headline)
            matched = True
            break

    # If no pattern matched, it's uncategorized
    if not matched:
        exclusive_category_results["Uncategorized"]["count"] += 1
        exclusive_category_results["Uncategorized"]["headlines"].append(headline)

# Calculate exclusive_sum properly
exclusive_sum = sum(res['count'] for res in exclusive_category_results.values())


# --- CO-OCCURRENCE HEATMAP (NEW IN V3) ---
print("\n" + "=" * 60)
print("GENERATING CO-OCCURRENCE HEATMAP...")
print("=" * 60)

# Create co-occurrence matrix
categories_for_heatmap = [cat for cat in patterns.keys()
                          if cat not in ["Uncategorized",
```

```
                                        "General 'Disability' Keyword"]]

co_occurrence_matrix = pd.DataFrame(0, index=categories_for_heatmap,
                                    columns=categories_for_heatmap)

# Count co-occurrences
for idx, row in df.iterrows():
    headline = str(row[headline_col])
    matched_cats = []

    for label in categories_for_heatmap:
        pattern = patterns[label]
        if re.search(pattern, headline):
            matched_cats.append(label)

    # Update co-occurrence matrix
    for i, cat1 in enumerate(matched_cats):
        for cat2 in matched_cats[i:]:
            co_occurrence_matrix.at[cat1, cat2] += 1
            if cat1 != cat2:
                co_occurrence_matrix.at[cat2, cat1] += 1

# Create heatmap
plt.figure(figsize=(14, 12))
sns.heatmap(co_occurrence_matrix, annot=True, fmt='d', cmap='YlOrRd',
            cbar_kws={'label': 'Number of Co-occurrences'},
            linewidths=0.5, linecolor='gray')
plt.title('Co-occurrence Matrix of Disability Categories in BBC Coverage\n'
          '(How often categories appear together)',
          fontsize=14, pad=20)
plt.xlabel('Category', fontsize=12)
plt.ylabel('Category', fontsize=12)
plt.xticks(rotation=45, ha='right')
plt.yticks(rotation=0)
plt.tight_layout()
plt.savefig('bbc_cooccurrence_heatmap_v3.png', dpi=300, bbox_inches='tight')
plt.show()


# --- VISIBILITY ANALYSIS ---
print("\n" + "=" * 60)
print("VISIBILITY ANALYSIS")
print("=" * 60)

# Define visibility categories
visible_categories = ["Deaf/Hearing", "Blind/Vision",
                      "Physical & Mobility", "Learning Disabilities"]
invisible_categories = ["Chronic Illness/Pain", "Mental Health & Neuro",
                        "Autism/Neurodiversity"]

# Calculate visibility ratios
visible_exclusive = sum(exclusive_category_results[cat]["count"]
                        for cat in visible_categories)
invisible_exclusive = sum(exclusive_category_results[cat]["count"]
                          for cat in invisible_categories)

visible_multi = sum(multi_category_results[cat]["count"]
                    for cat in visible_categories)
invisible_multi = sum(multi_category_results[cat]["count"]
                      for cat in invisible_categories)

```

```
print(f"\nExclusive Category Analysis:")
print(f"  Visible disabilities: {visible_exclusive} articles "
      f"({(visible_exclusive/len(df)*100):.1f}%)")
print(f"  Invisible disabilities: {invisible_exclusive} articles "
      f"({(invisible_exclusive/len(df)*100):.1f}%)")
print(f"  Ratio (visible:invisible): "
      f"{visible_exclusive/invisible_exclusive:.1f}:1")

print(f"\nMulti-Category Analysis:")
print(f"  Visible disabilities: {visible_multi} thematic instances")
print(f"  Invisible disabilities: {invisible_multi} thematic instances")
print(f"  Ratio (visible:invisible): {visible_multi/invisible_multi:.1f}:1")

print("\n" + "=" * 60)
print("Analysis complete!")
print(f"Multi-category: {multi_sum} matches across {len(df)} articles")
print(f"Exclusive: {exclusive_sum} articles uniquely categorised")
print(f"Average categories per article: {multi_sum/len(df):.2f}")
```

B Supplementary Tables

Table 2: Inter-rater Reliability for Manual Validation (Cohen's Kappa)

Category	κ
General Disability	0.92
Physical/Sensory	0.88
Invisible/Chronic	0.81
Institutional	0.85
Overall	**0.87**

Table 3: Co-occurrence Frequency Matrix (Top 10 Pairs)

Category Pair	**Co-occurrences**
SEND/Special Schools + Benefits/Care	20
Blind/Vision + Animals/Well-being	15
Benefits/Care + Family/Carer	10
Blind/Vision + Sports/Arts	9
Charity/Community + Sports/Arts	15
Physical/Mobility + Sports/Arts	7
SEND + Family/Carer	25
Deaf/Hearing + Sports/Arts	5
Accessibility + Infrastructure	5
Learning Disabilities + Family	3

C Example Headlines by Category

C.1 Visible Disabilities (High Coverage)

- **SEND/Special Schools (17.0%):** "New SEND school opens doors to 120 pupils", "Ofsted rates special school 'outstanding'", "Parents fight council over mainstream placement"

- **Blind/Vision (8.5%):** "Blind pianist wins international competition", "Guide dogs charity launches new campaign", "Braille technology transforms student's education"
- **Deaf/Hearing (5.9%):** "BSL interpreter makes history at festival", "Deaf awareness week highlights communication barriers", "Cochlear implant gives toddler gift of hearing"

C.2 Invisible Disabilities (Low Coverage)

- **Chronic Illness/Pain (3.5%):** "Living with ME: Every day is a battle", "Long Covid sufferer shares two-year journey", "Fibromyalgia warrior opens support centre"
- **Autism/Neurodiversity (1.6%):** "Autistic artist's work featured in gallery", "New autism-friendly shopping hours launched", "Neurodiversity celebration week begins"
- **Mental Health (0.8%):** "PTSD support group marks anniversary", "Student's anxiety journey inspires others", "Depression awareness campaign launches"

C.3 Uncategorised Headlines (5.5%)

Sample of headlines that did not match any category patterns:

- "'I had to watch Justin Timberlake from a car park'"
- "'With the boat and the stars, I'm unstoppable'"
- "'I play my violin like a tiny cello'"
- "'I've been threatened over my Please Offer Me A Seat badge'"
- "When it's time to face your toughest crowd yet"