

---

Review

# Identifying Health Care Services Offered in the HIV Care Continuum via a Machine Learning–Based Topic Modeling Approach: Exploratory Literature Review

---

SangA Lee<sup>1</sup>, PhD; Layoung Kim<sup>1,2</sup>, PhD; Mi-So Shim<sup>3</sup>, PhD; Gwang Suk Kim<sup>1</sup>, PhD

<sup>1</sup>Mo-Im Kim Nursing Research Institute, College of Nursing, Yonsei University, Seoul, Republic of Korea

<sup>2</sup>College of Nursing and Brain Korea 21 FOUR Project, Yonsei University, Seoul, Republic of Korea

<sup>3</sup>College of Nursing, Keimyung University, Daegu, Republic of Korea

## Corresponding Author:

Gwang Suk Kim, PhD

Mo-Im Kim Nursing Research Institute

College of Nursing, Yonsei University

50-1 Yonsei-ro, Seodaemun-gu

Seoul, 03722

Republic of Korea

Phone: 82 2-2228-3342

Fax: 82 2-2227-8303

Email: [gskim@yuhs.ac](mailto:gskim@yuhs.ac)

---

## Abstract

**Background:** It remains unclear whether the existing health care services reflect the HIV care continuum, which underscores the need for integrated care beyond viral suppression.

**Objective:** This study aimed to analyze the literature on health care services for people living with HIV to enhance the understanding of trends and knowledge structures.

**Methods:** A literature review was conducted using BERTopic, an advanced machine learning-based topic modeling technique. We searched PubMed, CINAHL, EMBASE, and Cochrane databases for English-language studies published between 2013 and 2023. Analyses were performed twice: first, to gain a broad understanding of the literature, and second, to examine the specific details of health care services described.

**Results:** Among the 11,269 articles screened, 204 studies met the inclusion criteria. Within the HIV care continuum, most studies focused on the treatment retention stage, while studies focusing on the long-term stage were limited. A broad literature analysis identified five key topics, with “ART adherence” emerging as the most prominent topic. A more comprehensive analysis of health care services within the literature revealed 7 topics, reflecting diverse delivery methods and content in providing health care services for people living with HIV. The predominant topic, “ART adherence and counseling,” encompassed the largest number of studies, indicating the strongest emphasis in the field. Notably, the distribution of topics exhibited a distinct pattern: while health care service diversity was the highest in the earlier stages of the HIV care continuum, it became increasingly limited in the later stages.

**Conclusions:** This study provides valuable insights into current HIV care services and highlights areas for future research and intervention. Despite the shift toward lifelong HIV management, existing literature remains heavily focused on medication treatment, overlooking the multifaceted health care needs of people living with HIV. Research disparities, particularly concerning vulnerable populations, underscore the need for more inclusive studies and tailored health care services. Efforts should be intensified to bridge these gaps, ensuring inclusive and equitable health care services across diverse populations and fostering interdisciplinary collaboration to meet the evolving needs of people living with HIV, thereby enhancing the HIV care continuum for all.

*JMIR Public Health Surveill* 2025;11:e65081; doi: [10.2196/65081](https://doi.org/10.2196/65081)

**Keywords:** HIV infections; continuity of patient care; health services; review; machine learning

## Introduction

Similar to other chronic illnesses, people living with HIV require a lifelong disease-specific regimen to maintain their physical, psychological, and social well-being [1]. While individuals with sustained viral suppression do not risk sexually transmitting the virus to others [2], those with poor psychological health and lower socioeconomic status—measured by factors such as income, housing, and education—are particularly vulnerable to dropping out of care. This vulnerability is exacerbated by HIV-specific barriers, including stigma, discrimination in health care settings, and individual risk factors [3-5]. To address these challenges, health care services must be scaled up to provide targeted support [3,4]. Moreover, as these barriers cannot be immediately eliminated, a more continuous and comprehensive approach to HIV treatment and life journey is necessary.

The global response to HIV treatment has focused on retaining people living with HIV in care to ensure their healthy lives and to end AIDS [6]. In 2013, the HIV care continuum was established to outline the steps for the continuity of treatment outcomes, namely: diagnosis, linkage to care, treatment initiation, treatment retention, and the achievement and maintenance of viral suppression [7,8]. Institutions such as the World Health Organization (WHO) [9], International Association of Providers of AIDS Care [10], and US Department of Health and Human Services [11] proposed various guidelines, including health care services for people living with HIV on the care continuum at the international and national levels. These guidelines have agreed on the importance of the HIV care continuum, which requires a broader scope of integrated HIV care to achieve holistic health beyond viral suppression [12].

In the context of continuity, the WHO highlights that integrated, people-centered care expands the capacity of patients to play a more active role in long-term survival [13]. Positioned as components of the HIV care continuum, health care services such as HIV testing, medication supply, antiretroviral therapy (ART) adherence monitoring, adverse event management, immunological and virological monitoring, and follow-up for missed visits have been explored in previous studies [14-16]. However, it remains unclear whether the services being offered to people living with HIV align with lifelong continuity of care. To provide effective health care services supporting the HIV care continuum, a comprehensive data review to gain an in-depth understanding of historical trends and key perspectives on the services being provided to people living with HIV must be conducted. There are global, regional, and national differences in coverage, HIV incidence, prevalence, and mortality [17]. Low-income countries have focused on high-priority services, such as child and maternal health, due to resource constraints and immediate health care needs; whereas high-income countries possess the infrastructure, funding, and policy frameworks necessary to support more advanced health care systems. In particular, high-income countries have prioritized the development of integrated care, a prominent policy theme aimed at addressing complex and interconnected health

issues, including chronic diseases and the HIV care continuum [18]. Furthermore, high-income countries often serve as benchmarks for global policy development, making their trends and outcomes particularly valuable for informing evaluation and service provision [19]. Their well-established systems, combined with consistent access to resources and comprehensive data, enable the study of integrated care systems in a way that minimizes variability and reduces potential confounding factors associated with differences in economic conditions or institutional capacities. This focus on high-income countries ensures greater comparability and facilitates the derivation of meaningful, context-specific insights. Therefore, health care services provided in high-income countries [20] need to be analyzed first to review the HIV care continuum from the long-term perspective.

As various HIV health care service studies are being conducted across treatment stages, individuals, and methods, the literature needs to be comprehensively reviewed. Traditional methods, such as manual literature reviews and qualitative content analysis, while valuable for deriving specific insights from research criteria, face significant limitations in handling large and complex datasets. These methods are time-intensive, rely heavily on human expertise, and are subjective to judgment, making them prone to inconsistency as well as bias [21,22]. Furthermore, these methods are less effective at detecting overarching trends, correlations, and topic evolution over time, particularly in diverse and extensive datasets. Manual keyword extraction and topic grouping often fail to fully capture nuanced relationships between concepts, limiting the ability to generate a comprehensive and graphical perspective of the literature. To review, summarize, classify, and detect trends more efficiently in large-scale documents, text analysis techniques offer a valuable and effective solution [23,24].

To address these challenges, this study used Bidirectional Encoder Representations from Transformers (BERT) developed by Devlin [25] in Google artificial intelligence (AI) Language. Among various text analysis technical models, BERT is recognized as one of the most recent and advanced machine learning-based approaches to topic modeling. Unlike traditional text analysis methods, techniques like BERT offer several advantages. Its ability to pretrain deep bidirectional representations enables BERT to capture the contextual meanings of words and sentences by considering both left and right contexts simultaneously [25], resulting in more accurate and nuanced topic modeling. An example of traditional methods is the Latent Dirichlet Allocation (LDA), which relies on simpler probabilistic frameworks and often fails to account for complex semantic relationships [25,26]. Studies comparing LDA and BERT have demonstrated that BERT outperforms LDA in extracting deeper and distinct topics, identifying intricate topic relationships in datasets with complex linguistic structures [27,28]. BERT is particularly effective at processing large datasets, efficiently summarizing, classifying, and detecting trends across a wide range of documents. Considering its strengths, it has been increasingly used in research, particularly literature reviews analyzing numerous studies as well as

social media analyses processing numerous tweets to extract insights on health care issues [29-32]. Its scalability reduces the time and resources required for manual analysis while ensuring consistency in topic extraction and classification. In addition, BERT excels in modeling topic correlations as well as their evolution over time, providing a dynamic and graphical perspective of the data that traditional methods struggle to achieve. Its ability to handle complex linguistic structures and low-resource tasks enables the identification of meaningful patterns in diverse contexts, further enhancing its utility for analyzing comprehensive datasets [33]. While BERT offers numerous strengths, it effectively addresses the limitations of traditional approaches by ensuring objectivity, scalability, and the ability to generate insights that would otherwise require extensive human effort and expertise. These capabilities make BERT particularly well-suited for reviewing the extensive and heterogeneous body of literature on HIV health care services. This study therefore used BERT to depict the bibliographic trends and knowledge structure of publications on health care services to improve the HIV care continuum for people living with HIV in high-income countries.

## Methods

### Design

This study used a quantitative content analysis using BERTopic modeling to examine the keywords and topics in the literature on health care services provided to people living with HIV. The methodological process comprised phases: (1) search strategy and selection, (2) data preprocessing, and (3) data analysis, using topic modeling with BERTopic.

### Search Strategy and Selection Process

Relevant literature was searched from PubMed, CINAHL, EMBASE, and Cochrane databases. Key search terms included Medical Subject Heading (MeSH) terms, trade words, and natural language words such as “care continuum,” “HIV/AIDS,” and “service.” The search was limited to studies published between January 2013—when the WHO first introduced comprehensive guidelines on ART for all populations across the HIV care continuum—and December 2023 [34]. The inclusion criteria were studies: (1) targeting adults diagnosed with HIV, (2) investigating or providing health care services (eg, interventions, services, or programs) to improve the HIV care continuum, and (3) any study design conducted in high-income countries. The exclusion criteria were: (1) studies targeting specific subpopulations (eg, drug users, individuals with mental illnesses, or sex workers), (2) studies targeting HIV prevention services, (3) studies for which the full text was unavailable in English, (4) studies targeting public policies, (5) literature reviews, and (6) duplicates. The exclusion of studies on specific subpopulations aimed to focus the review on general HIV care rather than focusing on tailored interventions for specific groups. While these populations face unique challenges and require specialized interventions, their exclusion was conducted to provide a broader perspective on the HIV care continuum.

This approach ensures that the analysis reflects the overall capacities and integration efforts of health care systems, offering insights applicable to the wider population rather than a subset of individuals.

In total, 14,377 articles were retrieved from the databases. After removing duplicates, 11,269 articles were screened for their titles, abstracts, and full texts to determine their eligibility. Multiple studies using the same health care service were assessed through text frequency analysis to detect keyword matches. Most exhibited few keyword matches, suggesting distinct studies. Subsequently, all studies, with the exception of secondary analyses, underwent a full-text review. Three researchers (SL, MS, and LK) independently screened all the articles and selected the final list of articles for analysis after discussion. Discrepancies were discussed with the principal investigator (GK) and consensus was reached as to whether to include the article. The dataset size selected was deemed appropriate for BERT, as it demonstrated efficacy in handling modest-sized datasets with a comparable scale in previous studies, often using 100 documents [28,35]. As with many bibliometric analysis studies exploring research trends [36], a quality assessment was unnecessary. All the data, including the metadata elements of the title, abstract, authors, and year, were downloaded for analysis.

### Data Preprocessing

Before data analysis, a sequence of preprocessing steps were conducted to clean and prepare the data for further analysis. Articles without abstracts were excluded to ensure a comprehensive set of textual information. Additional columns were created to classify studies by country, HIV care continuum, and health care services provided. The countries from which each study collected data and the specific stages of the HIV care continuum on which each study focused were identified to derive more meaningful interpretations of the distribution of research topics. Specific stages of the HIV care continuum, including linkage to care, treatment initiation, treatment retention, and the long-term stage [7,8], were identified. The classification was finalized following a consensus reached through discussion between 2 researchers (SL and LK), with multiple stages being investigated in one study. After incorporating studies covering multiple stages, the final corpus comprised 242 documents for analysis. To ensure a more thorough investigation of health care services, detailed descriptions of services from the methods sections of each article were extracted. This process, conducted by 2 researchers (SL and LK), enhanced the accuracy of the dataset. The usage of details regarding the health care services described in the methods section helped create a more robust text corpus, enhancing the efficacy of the subsequent topic modeling analysis.

Stop words were then excluded, and a replacement dictionary was generated. Stop words, including “a,” “the,” and “of,” refer to common words that appear frequently in the text but do not convey insights into the specific topic of a document. All English stop words were removed using the Natural Language Toolkit library in Python. Words

that barely contained significant meanings for the research purpose and degraded or hindered the interpretability of results (eg, background, method, results, and trial registration) were added to the toolkit. The extended stop words were then excluded from the data. The replacement dictionary was used to combine several words or abbreviations with the same or similar meanings and designate a single representative word (eg, people living with HIV, HIV, and AIDS). During this stage, repeated word analyses were performed with regard to abstract, and the researchers reached consensus on the words to be registered in the dictionaries through discussion.

## Data Analysis

BERTopic, a Python-based package, was used to extract the primary topics from the literature. It was implemented with the all-MiniLM-L6-v2 embedding model to ensure accurate and efficient document embedding, leveraging its compact yet powerful transformer architecture to generate high-quality sentence representations for topic modeling [37,38]. The BERTopic process consists of three steps: (1) document embedding, (2) document clustering, and (3) topic representation [39,40]. Document embedding, using the all-MiniLM-L6-v2, reads sentences in both directions, pretrains them in deep representations, and then generates word embeddings (ie, representations of valued vectors transformed from words to numbers depending on the surrounding contexts). Document clustering applies Uniform Manifold Approximation and Projection to reduce the dimensions [41] and cluster semantically similar words through the Hierarchical Density-Based Spatial Clustering of Applications with the Noise algorithm [42]. Topic representation extracts significant words on class-based term frequency-inverse document frequency, which is a metric to measure the relevance of words [43].

To validate the findings, LDA—a widely used topic modeling method—was also applied. Since no definitive method exists to determine the number of topics, the research team conducted multiple discussions to assess topics generated by the BERT-based approach. The final number of topics was determined by considering evaluation metrics, intertopic distance maps, data characteristics, and interpretability of the results, in comparison to LDA [25,44,45]. The evaluation metrics used to assess topic quality were topic coherence and topic diversity, both ranging from 0 to 1. Topic coherence measures how frequently words co-occur in a dataset, with higher values indicating more interpretable topics, while topic diversity assesses the uniqueness of words within topics, with higher values indicating more distinct topics [46]. The intertopic distance map visualizes topics as circles in a 2D space, where the size of each

circle represents the number of studies assigned to that topic, and the proximity between circles indicates the degree of similarity between topics [47]. In this study, the BERT-generated model demonstrated superior evaluation metrics and a more balanced distribution of topics in the intertopic distance maps, with minimal overlap. In addition, the BERT approach yielded more data-related, distinct, and plausible results, outperforming LDA topic modeling in both structure and longitudinal topic trends. In line with the common practice in prior studies using BERTopic, Topic -1—a nonthematic cluster of documents with weak thematic alignment and often containing words commonly found across different topics [48]—was excluded from the interpretation of results. The proportion of nonthematic documents assigned to Topic -1 varied among the studies, ranging from approximately zero to 54% [25,36]. In this study, the analyses were performed twice, with one aimed at broadening the understanding of the literature related to health care services for people living with HIV and the other focused on a specific exploration of the health care services described in the literature.

## Ethical Considerations

The study was exempted from the approval process because it was a review and analysis of existing literature. No human participants were not directly involved, and no personal or private information was used.

## Results

### Data Description

A total of 204 studies were included in the analysis (Figure 1). Over the past decade, an average of approximately 19 studies on health care services for people living with HIV were conducted annually, ranging from 10 to 31 studies (Figure 2). The number of publications increased significantly between 2018 and 2020, but declined in 2016 and from 2021 onwards. Table 1 summarizes the characteristics of the included studies. The majority of studies were conducted in the United States (n=174, 85.3%). Regionally, as defined by the WHO [49], the Region of the Americas (AMR) accounted for the highest representation, while the Eastern Mediterranean Region was represented by only one study. The average age of study participants was 40.52 (SD 10.67) years, with the longest recorded duration since their HIV diagnosis being 30 years. Among the HIV care continuum stages, the treatment retention stage comprised the highest number of studies, totaling 185, while the long-term stage had the fewest studies.

Figure 1. PRISMA flowchart for article selection.

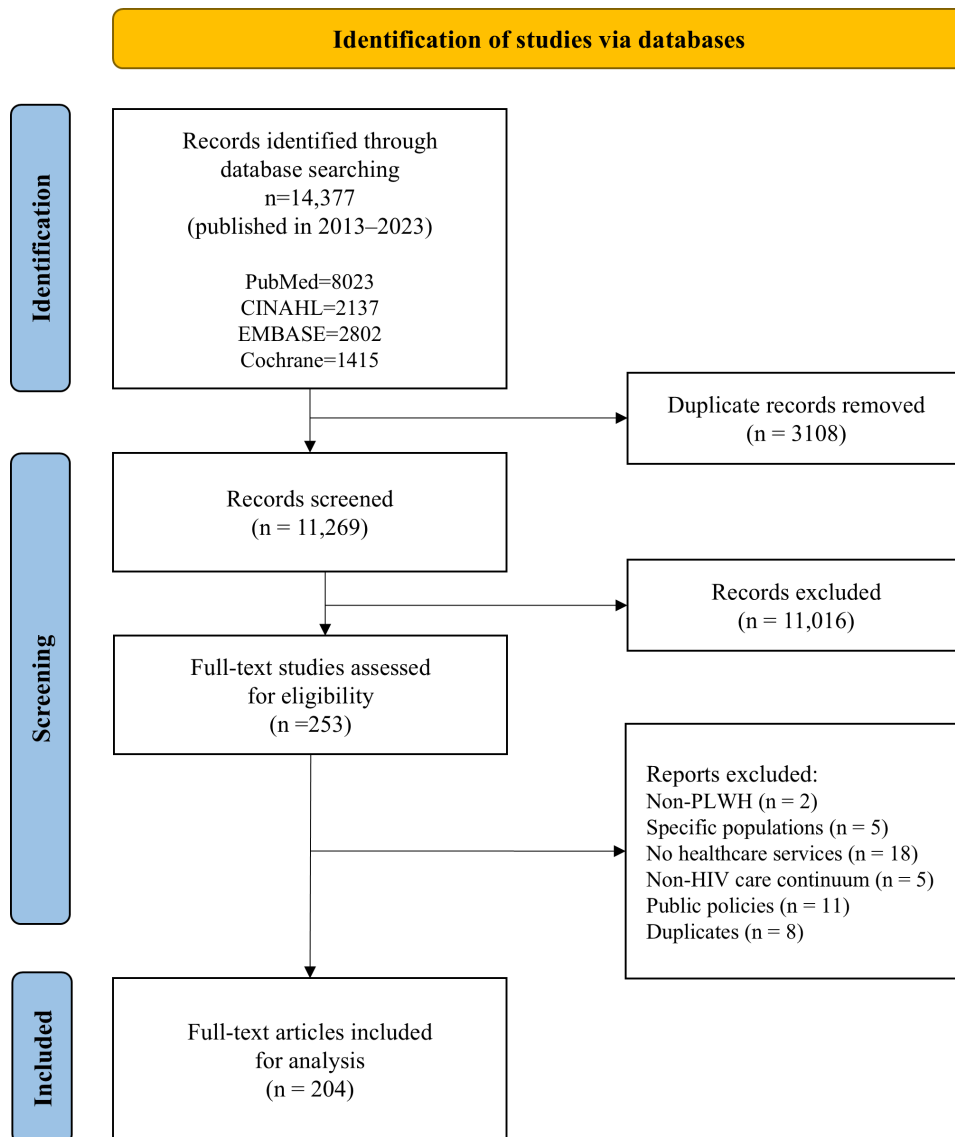
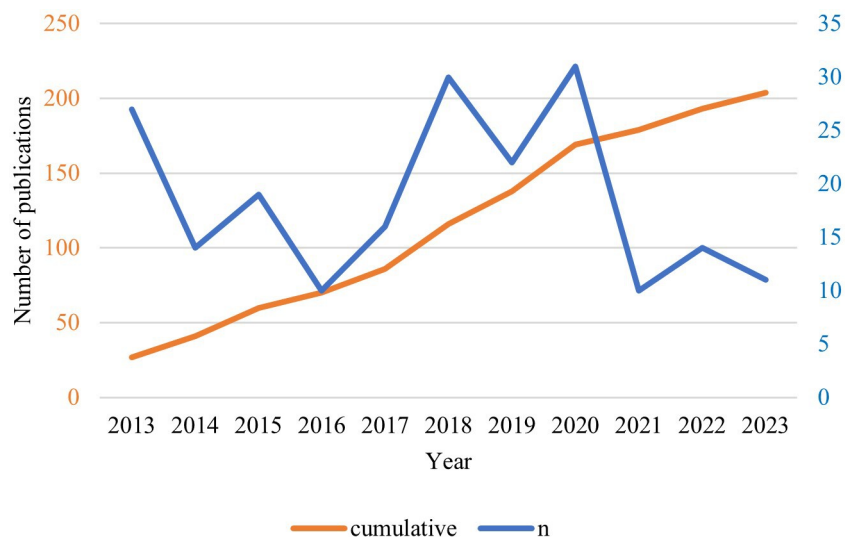


Figure 2. Number of publications on health care services for people living with HIV from 2013 to 2023.



**Table 1.** Characteristics of the analyzed studies (N=204).

Characteristics	Values
Research design, n (%)	
Quantitative	164 (80.4)
Qualitative	22 (10.8)
Mixed methods	18 (8.8)
Sample	
Size, min-max	10-32,999
Age, years	18-85
Period of HIV diagnosis, years	0-30
Country, n (%)	
Region of the Americas, n (%)	
United States	174 (85.3)
Canada	6 (2.9)
European Region, n (%)	
Spain	8 (3.9)
United Kingdom	4 (2.0)
Denmark	1 (0.5)
Italy	1 (0.5)
Netherlands	1 (0.5)
Poland	1 (0.5)
Switzerland	1 (0.5)
Western Pacific Region, n (%)	
Australia	3 (1.5)
Korea	2 (1.0)
Singapore	1 (0.5)
Oman	1 (0.5)
HIV care continuum <sup>a</sup> , %	
Linkage to care	17
Treatment initiation	29
Treatment retention	185
Long-term	11

<sup>a</sup>Records duplicates.

## Topic Modeling on Health Care Services Research for People Living With HIV

BERTopic was initially applied to the abstract corpus to identify overarching trends and the knowledge structure within the included publications. Among the 242 documents, 191 were classified into 5 distinct topics, while the remaining 51 (21.1%) were identified as a nonthematic topic. This proportion is relatively small compared to previous studies, where the proportion ranged from 0% to 54% [25,36]. The most frequently occurring keywords within the nonthematic topic included “hiv,” “intervention,” “health,” “use,” and “improve,” which represent broad and general concepts rather than distinct thematic categories.

Table 2 presents the 5 distinct topics, labelled by the research team, alongside their respective keywords and the number of documents assigned to each topic. The 2 prominent topics—“ART adherence” and “Retention and engagement in care”—align with key stages of the HIV care continuum. The remaining 3 topics focus on specific populations: “Minorities-focused” consists of studies concentrating on race and ethnicities, while “Young population-focused,” emerged as an age-specific category. In addition, “Newly diagnosed men who have sex with men (MSM)” encompasses health care services tailored to a specific population at a specific stage.

**Table 2.** Topics and top keywords identified in health care services research for people living with HIV.

Topics	Top keywords (importance weights)	n
ART <sup>a</sup> adherence	adherence (0.0470), intervention (0.0274), hiv (0.0267), art (0.0255), use (0.0229), medication (0.0221), group (0.0184), people living with HIV (0.0172), improve (0.0164), self (0.0158)	120 (62.8%)
Retention and engagement in care	retention (0.0575), hiv (0.0505), intervention (0.0344), clinic (0.0253), health (0.0226), peer (0.0221), program (0.0209), improve (0.0208), adult (0.0191), engagement (0.0175)	29 (15.2%)

Topics	Top keywords (importance weights)	n
Minorities-focused	hiv (0.0451), component (0.0357), intervention (0.0292), black (0.0251), interview (0.0241), american (0.0237), african (0.0219), african-american (0.0219), use (0.0214), latino (0.0209)	16 (8.4%)
Young population-focused	intervention (0.0641), adherence (0.0586), youth (0.0504), adolescent (0.0344), art (0.0342), hiv (0.0336), condition (0.0292), accept (0.0261), medication (0.0246), group (0.0240)	14 (7.3%)
Newly diagnosed MSM <sup>b</sup>	test (0.0572), hiv (0.0496), ltc (0.0428), clinic (0.0387), client (0.0341), msm (0.0303), diagnose (0.0301), hiv-test (0.0250), case (0.0234), men (0.0216)	12 (6.3%)

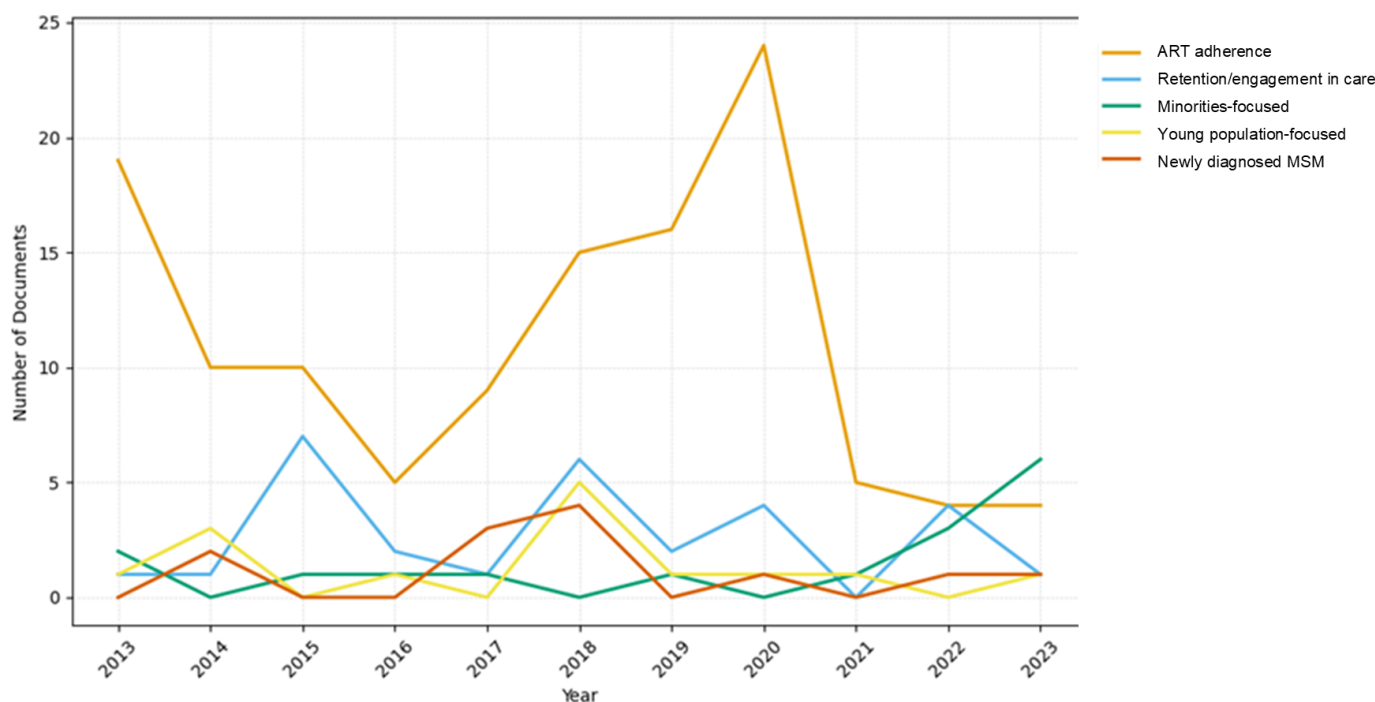
<sup>a</sup>ART: antiretroviral therapy.

<sup>b</sup>MSM: men who have sex with men.

When comparing the evaluation metrics of the models generated using LDA and BERTopic, the topic coherence scores were 0.2987 and 0.4264, respectively, while the topic diversity scores were 0.4200 and 0.6667, respectively, indicating superior performance of BERTopic in both aspects. In addition, the intertopic distance map of the BERT-generated model presented a more well-balanced distribution of topics with no overlap, whereas the LDA-generated model exhibited substantial topic overlap ([Multimedia Appendix 1](#)). Considering these comparisons, including topic interpretability, the BERTopic-generated model was determined to be the optimal choice for analysis.

[Figure 3](#) depicts the distribution of topics over the years. The research focus on “ART adherence” peaked in 2020, while notable declines were observed during mid-2010 and after 2020. “Retention and engagement in care” exhibited fluctuating trends. Coinciding with the onset of the COVID-19 pandemic, studies involving people living with HIV tended to decrease post 2020, but research on “Minorities-focused” topic exhibited recent increases.

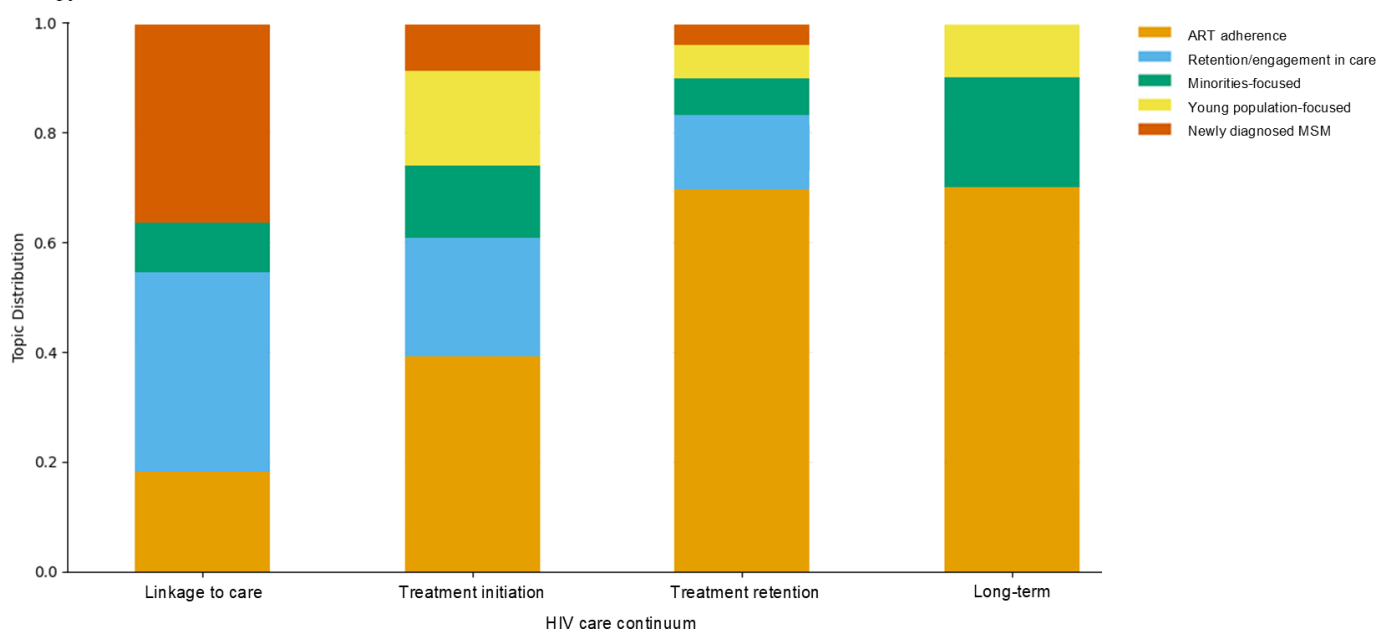
**Figure 3.** Topic distribution in health care services research for people living with HIV from 2013 to 2023. ART: antiretroviral therapy.



Among the 4 stages of the HIV care continuum, the treatment initiation and retention stages were investigated across all topics ([Figure 4](#)). In the first stage (linkage to care), research on young populations was notably lacking, whereas the “Newly diagnosed MSM” topic had the highest representation compared to other stages. In the treatment initiation stage, the “Young population-focused” topic began to emerge, with all topics being more evenly distributed. The

“ART adherence” topic accounted for more than half of the studies in the treatment retention stage. Finally, the long-term stage had the fewest studies overall. Regarding the regional distribution of topics by country, only the AMR accounted for all topics, while studies from other regions were limited. Research conducted in the AMR predominantly focused on ART adherence.

**Figure 4.** Distribution of topics in health care services research for people living with HIV across the HIV care continuum. ART: antiretroviral therapy; MSM: men who have sex with men.



### Topic Modeling on Health Care Services for People Living With HIV

The second analysis aimed to provide more comprehensive examination of health care services by using the detailed descriptions of health care services found in the methods section of the articles. When running BERTopic, 181 out of the total 242 documents were categorized into 7 topics, while the remaining 61 (25.2%) fell into a nonthematic category. The most frequent occurring keywords in the nonthematic topic included “care,” “participant,” “hiv,” “patient,” and “health,” suggesting that these documents contained broad, general concepts rather than distinct thematic distinctions.

Table 3 presents the 7 topics, as labeled by the research team, along with their respective keywords and the number of documents assigned to each. Among these, “ART adherence and counseling” was the most prevalent topic, reflecting strategies to enhance ART adherence, as indicated by the

keywords “counseling” and “message.” Other topics explore more specific aspects of health care service delivery models, target populations, and service content. The “Peer support” and “Pharmacist involvement” topics highlight the engagement of both health care and non-health care professionals in studies providing health care services to people living with HIV. The “Pharmacist involvement” and “Youth-to-adults transition care” topics specify who delivers health care services and for whom, respectively. “Monetary incentives” pertain to studies using financial rewards as benefits for people living with HIV. The topic “Health promotion and care navigation” focuses on structured interventions designed to guide individuals in maintaining and improving health. Finally, the “Linkage to care” represents studies addressing a specific stage of the HIV care continuum. As reflected in their labels, the 7 topics collectively reflect a combination of delivery methods and content in the provision of health care services to people living with HIV.

**Table 3.** Topics and top keywords identified as health care services for people living with HIV.

Topics	Top keywords (importance weights)	n
ART <sup>a</sup> adherence and counseling	adherence (0.0312), participant (0.0283), session (0.0280), medication (0.0212), intervention (0.0209), use (0.0186), patient (0.0184), message (0.0150), counsel (0.0139), receive (0.0128)	81 (44.8%)
Peer support	care (0.0369), hiv (0.0344), peer (0.0246), health (0.0229), intervention (0.0211), support (0.0196), participant (0.0190), patient (0.0182), session (0.0176), barrier (0.0172)	40 (22.1%)
Youth-to-adults transition care	care (0.0572), hiv (0.0473), patient (0.0435), team (0.0329), clinic (0.0276), transition (0.0253), visit (0.0222), service (0.0213), provide (0.0200), adult (0.0183)	20 (11.0%)
Pharmacist involvement	art (0.0606), patient (0.0568), medication (0.0389), pharmacist (0.0382), hiv (0.0375), clinic (0.0304), day (0.0292), drug (0.0283), pharmd (0.0246), hiv-pharmd (0.0246)	17 (9.4%)
Monetary incentives	chance (0.0482), participant (0.0449), point (0.0440), chance-win (0.0437), win (0.0437), hiv (0.0416), component (0.0395), prize (0.0351), tm (0.0328), test (0.0320)	9 (5.0%)
Health promotion and care navigation	hiv-care (0.0662), care (0.0613), hiv (0.0611), health (0.0514), health-promotion (0.0493), promotion (0.0490), care-navigation (0.0480), message (0.0479), navigation (0.0435), participant (0.0434)	8 (4.4%)



Topics	Top keywords (importance weights)	n
Linkage to care	care (0.0684), hiv (0.0554), site (0.0545), people living with HIV (0.0407), hiv-care (0.0347), classify (0.0341), ltc (0.0336), ed (0.0312), program (0.0311), patient (0.0289)	6 (3.3%)

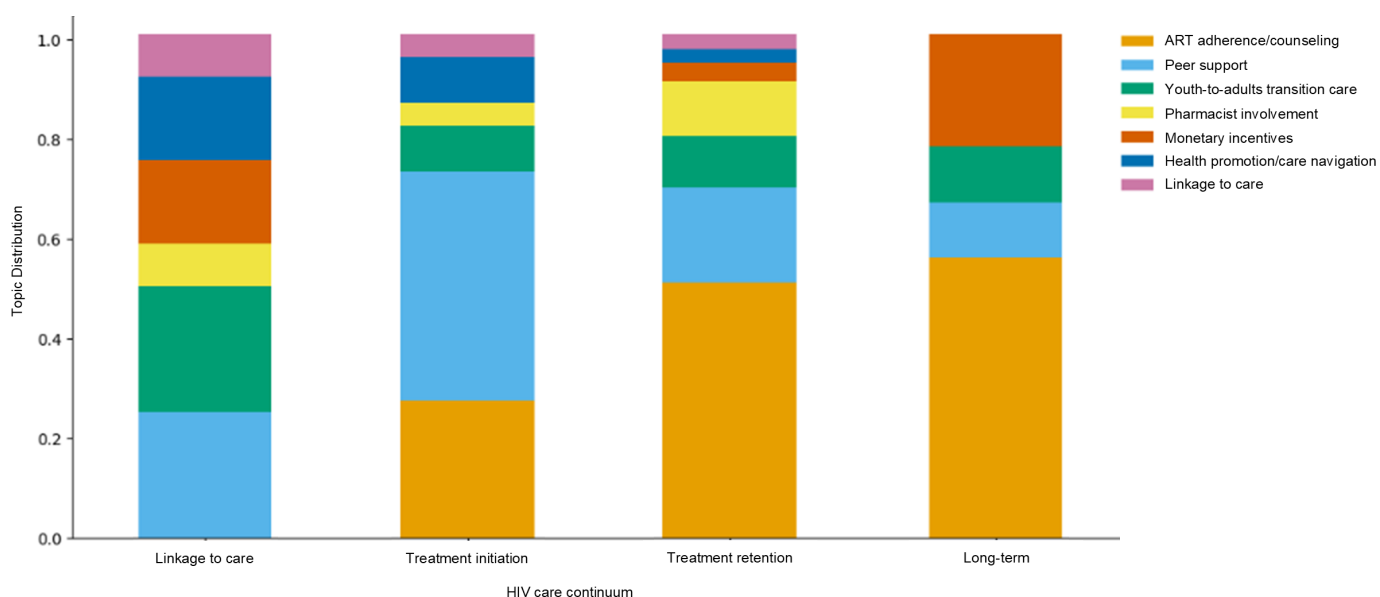
<sup>a</sup>ART: antiretroviral therapy.

When comparing the evaluation metrics of the models generated using LDA and BERTopic, the topic coherence scores were 0.3799 and 0.4458, respectively, while the topic diversity scores were 0.4429 and 0.6125, respectively, demonstrating BERTopic's superior performance in both measures. Furthermore, the intertopic distance map of the BERT-generated model exhibited a well-balanced distribution of topics with no overlap, whereas the LDA-generated model displayed overlapping topics (Supplementary Figure 1 in [Multimedia Appendix 1](#)). Considering these factors, including topic interpretability, the BERTopic-generated model outperformed the LDA-generated model and was ultimately selected as the preferred approach.

Figure 5 illustrates the distribution of health care service topics across each stage of the HIV care continuum, highlighting varying levels of research coverage. While research on the initial linkage to care and final long-term

stages were similarly scarce, the distribution of health care service topics presented distinct patterns. At the linkage to care stage, all health care services were represented except those related to medication adherence. In the treatment initiation stage, services focusing on medication adherence began to emerge, with peer support being the most prominent. During the treatment retention stage, medication adherence-focused services accounted for more than half of the studies. In the long-term stage, services related to medication adherence and monetary incentives became more prominent, whereas services involving pharmacist involvement and health navigation, which were present in earlier stages, declined. Overall, the diversity of health care services was the highest in the earlier stages of the continuum but became less varied in later stages, where research was more limited.

**Figure 5.** Distribution of health care service topics for people living with HIV across the HIV care continuum. ART: antiretroviral therapy; MSM: men who have sex with men.



## Discussion

### Principal Findings

The HIV care continuum is a framework that describes the treatment journey of people living with HIV, encompassing the stages from diagnosis to linkage to care, engagement in care, and viral suppression through ART [7]. Key health care service components within this continuum include HIV testing, medication supply, treatment adherence monitoring, adverse event management, as well as virological monitoring [14-16]. However, like other chronic diseases, the HIV care continuum requires a shift toward an integrated care system centered on patient self-management, equipping individuals

with the health care resources needed to manage their condition throughout life [50]. Moreover, as patients are increasingly expected to take active roles in managing their success of HIV treatment and long-term health [51], it is critical to move beyond the sole objective of viral suppression and emphasize overall quality of care for all people living with HIV [12]. Nonetheless, the extent to which current services align with the principles of lifelong continuity of care remains unclear. To address this gap, this study aimed to enhance the understanding of health care service trends and knowledge structures within the HIV care continuum by reviewing the literature using BERT. Given that approximately a decade has passed since the global recognition of the HIV care continuum, reviewing research in this field

is essential for identifying both dominant areas and those lacking attention and support.

In this study, research on health care services within the HIV care continuum has primarily focused on the treatment retention stage, with a notable emphasis on ART adherence. While the importance of the HIV care continuum is widely recognized [12], research beyond disease-focused models remains underexplored in the field. Notably, despite longstanding concerns regarding the quality of life (QoL) of people living with HIV—including issues such as comorbidities, stigma, depression, discrimination, and sexual health [52]—these topics were not prominently represented in the extracted themes. The dominance of treatment-focused topics reveals persistent gaps in the literature, even amid global calls for more comprehensive, continuum-based HIV care. To broaden perspectives on life beyond treatment, future research should embrace emerging concepts such as “long-term success” [53] and “healthy living” [54] for people living with HIV, which emphasize a holistic approach addressing the multifaceted needs of people living with HIV. Optimizing health-related QoL and addressing age-related concerns is crucial to ensuring people living with HIV can lead fulfilling lives. This highlights the importance of multidisciplinary teams in delivering integrated, person-centered care and support for people living with HIV.

From a population perspective, there remains a notable lack of research on health care services tailored to minority populations. The topic of MSM emerged as a distinct research focus, reflecting the longstanding dominance of this demographic among people living with HIV. Consequently, research targeting MSM has been abundant. However, research addressing racial and ethnic minorities was consolidated into a single topic, and no distinct topic related to older adults emerged. This gap may be attributable to challenges in engaging these populations, particularly during the COVID-19 pandemic. Such academic trends inadequately address the global HIV epidemic. In high-income countries such as the US, HIV prevalence and incidence estimates are disproportionately higher among Black or African American and Hispanic or Latino individuals compared to White individuals. In addition, women are disproportionately affected among acquiring HIV through heterosexual contact, highlighting their vulnerability despite lower numerical representation [55]. The disease burden attributable to HIV infection among older adults, including prevalence, mortality, and QoL concerns, has increased [56]. These data underscore the critical need for comprehensive actions across the HIV care continuum and increased support tailored to younger and older populations, irrespective of race, ethnicity, and gender. Prioritizing and funding research aimed at understanding the specific needs of diverse and vulnerable populations is critical to fostering a more equitable health care landscape for HIV-infected individuals.

The topics identified on health care services aligned well with the needs of people living with HIV. Peer support and youth-to-adult transition services were prominent in the early stages of the HIV care continuum, when people living with HIV struggle with limited information following diagnosis

and must decide whether to undergo ART. This period also represents a high-risk time for disengagement, particularly among young adults with HIV transitioning from pediatric to adult-oriented care [57,58]. ART adherence and counseling services were primarily implemented in the subsequent stages, supporting people living with HIV as they continued treatment and worked toward clinical stability [8]. While the health care services identified in the literature corresponded to different stages of the HIV care continuum, they were predominantly concentrated in the treatment retention stage, with fewer and less diverse services available in other stages. Therefore, more health care services should be developed and implemented across the HIV care continuum, particularly at the long-term stage. Specifically, more varied services to facilitate treatment initiation for people living with HIV at the initial stage are necessary, along with supportive services to promote healthy living and aging with HIV in the long term [59]. Furthermore, the cost-effectiveness, feasibility, sustainability, and confidentiality of these services must be further investigated, particularly for marginalized, younger, and older people living with HIV [60,61].

Pharmacists emerged as a distinct topic in health care services, whereas nurses and physicians were not separately identified. This may be attributed to the unique nature of HIV infection, which heavily relies on ART. Since the introduction of the first ART drug in 1987, ART has been continuously and actively developed and evaluated [62]. Furthermore, the WHO has continuously issued updated evidence-based ART guidelines since publishing its first consolidated guidelines in 2013 [33]. Recent trends have emphasized the cost-effectiveness and accessibility of ART, alongside advancements in potential cures and vaccines [62]. Given the dependency on ART and active trends in its advancement, pharmacies serve as crucial locations for the delivery of HIV treatment, testing, and prevention services, although the scope of services varies among countries. Pharmacists' attitudes and services that are favorable to people living with HIV are essential for enhancing the HIV care continuum [63]. Expanding the services offered by pharmacies and pharmacists as primary service providers for testing and medication, particularly for high-risk and vulnerable populations, should be key consideration [64].

Although physicians and nurses play fundamental roles in HIV care, addressing evolving needs for health and longevity requires extending care beyond primary care to optimal long-term management beyond treatment. This need is particularly pertinent in high-income countries with abundant resources [65,66]. Nurses' practices have been closely aligned with the needs of people living with HIV in HIV care, contributing to interdisciplinary efforts that facilitate the active development and implementation of interventions across the HIV care continuum [67]. However, their involvement in policymaking at local, national, and international levels remains limited [67]. Given their deep understanding of the needs of people living with HIV, nursing professionals should actively participate in decision-making processes to reshape practices and policies within the HIV care continuum. Furthermore, while not a distinct topic in this

study, social welfare professionals also play a crucial role in the long-term management of HIV.

In addition, involving family members in health care services is recommended to optimize the HIV care continuum, particularly for key populations such as women and adolescents [10]. Numerous studies advocate a holistic patient-centered approach for people living with HIV, which involves not only pharmacists and peers, but also partners, family, and other individuals. Ensuring a comprehensive approach that encompasses a diverse range of health care and non-health care professionals over the long term is critical for promoting the HIV care continuum, particularly for vulnerable populations.

Overall, a decline in the number of studies on HIV services was observed, likely influenced by the impact of the COVID-19 pandemic. Notable drops in research output occurred between 2014 and 2016, and again in 2021, coinciding with global outbreaks of infectious diseases such as Ebola, Zika, and COVID-19 [68]. Similar to the upwards trend observed after the resolution of outbreaks such as Ebola and Zika, research efforts may increase shortly after the end of the COVID-19 pandemic. Therefore, research published from this perspective is important. As the vast majority of HIV research is conducted in the US, findings from this context may not be generalizable to other countries owing to significant sociocultural differences. Expanding research efforts across diverse countries and tailoring health care services to the specific needs of each country is crucial.

This study has several limitations. First, it focused solely on studies from high-income countries; studies from middle- and low-income countries may present different characteristics. Second, studies that did not explicitly address the HIV care continuum were excluded, necessitating careful interpretation of the findings. Finally, our trend analysis relied on linear models, potentially oversimplifying the intricate evolution of topics influenced by multiple factors.

## Conclusions

Despite the evolving nature of HIV care toward life-long management, the literature predominantly emphasizes medical treatment, highlighting significant gaps in addressing the multifaceted needs of people living with HIV. This study provides a novel contribution by analyzing health care services aimed at enhancing HIV care continuum for people living with HIV using BERT, an advanced machine learning-based topic modeling method. By examining research trends across different stages of the HIV care continuum, this study offers valuable insights into both overarching research trends and specific service-related themes. Notably, it extends beyond traditional treatment-focused disease models to encompass long-term management and holistic well-being considerations, underscoring the importance of interdisciplinary collaboration in meeting the evolving needs of people living with HIV. Overall, this research advances understanding of HIV care and provides critical directions for future interventions and studies in this field.

## Acknowledgments

This work was supported by the National Research Foundation of Korea funded by the Korean government (Ministry of Science and ICT) [Grant no. 2020R1A2C101081713], the Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Education [Grant no. 2020R1A6A1A03041989], and a Faculty Research Grant from Yonsei University College of Nursing [Grant no. 6-2022-0034]. Layoung Kim received a scholarship from the Brain Korea 21 FOUR Project, funded by the National Research Foundation of Korea, Yonsei University College of Nursing.

## Data Availability

Data sharing is not applicable to this article as no data sets were generated or analyzed during this study.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Supplementary Figure 1. Comparison of inter-topic distance maps generated utilizing BERTopic and LDA. [[PNG File \(Portable Network Graphics File\), 923 KB-Multimedia Appendix 1](#)]

## Checklist 1

PRISMA-ScR (Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews) Checklist.

[[DOCX File \(Microsoft Word File\), 110 KB-Checklist 1](#)]

## References

1. Swendeman D, Ingram BL, Rotheram-Borus MJ. Common elements in self-management of HIV and other chronic illnesses: an integrative framework. *AIDS Care*. Oct 2009;21(10):1321-1334. [doi: [10.1080/09540120902803158](https://doi.org/10.1080/09540120902803158)] [Medline: [20024709](https://pubmed.ncbi.nlm.nih.gov/20024709/)]
2. The Lancet. U=U taking off in 2017. *Lancet HIV*. Nov 2017;4(11):e475. [doi: [10.1016/S2352-3018\(17\)30183-2](https://doi.org/10.1016/S2352-3018(17)30183-2)] [Medline: [29096785](https://pubmed.ncbi.nlm.nih.gov/29096785/)]

3. Choi JP, Seo BK. HIV-related stigma reduction in the era of undetectable equals untransmittable: the South Korean perspective. *Infect Chemother*. Dec 2021;53(4):661-675. [doi: [10.3947/ic.2021.0127](https://doi.org/10.3947/ic.2021.0127)] [Medline: [34979602](https://pubmed.ncbi.nlm.nih.gov/34979602/)]
4. Safreed-Harmon K, Anderson J, Azzopardi-Muscat N, et al. Reorienting health systems to care for people with HIV beyond viral suppression. *Lancet HIV*. Dec 2019;6(12):e869-e877. [doi: [10.1016/S2352-3018\(19\)30334-0](https://doi.org/10.1016/S2352-3018(19)30334-0)] [Medline: [31776099](https://pubmed.ncbi.nlm.nih.gov/31776099/)]
5. Moen M, Doede M, Johantgen M, et al. Nurse-led hospital-to-community care, clinical outcomes for people living with HIV and health-related social needs. *J Adv Nurs*. May 2023;79(5):1949-1958. [doi: [10.1111/jan.15485](https://doi.org/10.1111/jan.15485)] [Medline: [36345144](https://pubmed.ncbi.nlm.nih.gov/36345144/)]
6. Joint United Nations Programme on HIV/AIDS. Prevailing against pandemics by putting people at the centre: world AIDS day report 2020. Joint United Nations Programme on HIV/AIDS; 2020. URL: <https://www.unaids.org/en/resources/documents/2020/prevailing-against-pandemics> [Accessed 2024-03-27]
7. Babalola S, Van Lith LM, Mallalieu EC, et al. A Framework for health communication across the HIV treatment continuum. *J Acquir Immune Defic Syndr*. Jan 1, 2017;74(Suppl 1):S5-S14. [doi: [10.1097/QAI.0000000000001206](https://doi.org/10.1097/QAI.0000000000001206)] [Medline: [27930606](https://pubmed.ncbi.nlm.nih.gov/27930606/)]
8. Kim GS, Kim Y, Park MK, Lee SA, Lee Y, Shim MS. Barriers and facilitators on the HIV care continuum among adults living with HIV in high-income countries: a scoping review protocol. *Syst Rev*. Oct 15, 2022;11(1):222. [doi: [10.1186/s13643-022-02097-x](https://doi.org/10.1186/s13643-022-02097-x)] [Medline: [36243736](https://pubmed.ncbi.nlm.nih.gov/36243736/)]
9. World Health Organization. HIV Treatment: Guidelines for Managing Advanced HIV Disease and Rapid Initiation of Antiretroviral Therapy. Policy Brief; 2017. URL: <https://www.who.int/publications/i/item/9789241550062> [Accessed 2024-03-27]
10. International Advisory Panel on HIV Care Continuum Optimization. IAPAC guidelines for optimizing the HIV care continuum for adults and adolescents. *J Int Assoc Provide AIDS Care*. 2015;14(1):S3-S34. [doi: [10.3389/10.1177/2325957415613442](https://doi.org/10.3389/10.1177/2325957415613442)] [Medline: [26527218](https://pubmed.ncbi.nlm.nih.gov/26527218/)]
11. The White House. National HIV/AIDS strategy for the united states 2022–2025. 2021. URL: <https://files.hiv.gov/s3fs-public/NHAS-2022-2025.pdf> [Accessed 2024-03-27]
12. Horberg MA, Certa JM, Rubenstein KB, et al. Beyond the HIV care continuum and viral suppression: broadening the scope of quality metrics for total HIV patient care. *AIDS Patient Care STDS*. Nov 2020;34(11):461-469. [doi: [10.1089/apc.2020.0145](https://doi.org/10.1089/apc.2020.0145)] [Medline: [33147087](https://pubmed.ncbi.nlm.nih.gov/33147087/)]
13. World Health Organization. Continuity and Coordination of Care: A Practice Brief to Support Implementation of the WHO Framework on Integrated People-Centred Health Services (9241514035). 2018. URL: <https://www.who.int/publications/i/item/9789241514033> [Accessed 2024-03-27]
14. McNulty MC, Schneider JA. Care continuum entry interventions: seek and test strategies to engage persons most impacted by HIV within the United States. *AIDS*. Feb 20, 2018;32(4):407-417. [doi: [10.1097/QAD.0000000000001733](https://doi.org/10.1097/QAD.0000000000001733)] [Medline: [29381558](https://pubmed.ncbi.nlm.nih.gov/29381558/)]
15. Mugavero MJ. Elements of the HIV care continuum: improving engagement and retention in care. *Top Antivir Med*. 2016;24(3):115-119. [Medline: [27841983](https://pubmed.ncbi.nlm.nih.gov/27841983/)]
16. Pinto RM, Witte SS, Filippone PL, Baird KL, Whitman WR. Factors that influence linkages to HIV continuum of care services: implications for multi-level interventions. *Int J Environ Res Public Health*. Nov 7, 2017;14(11):1355. [doi: [10.3390/ijerph14111355](https://doi.org/10.3390/ijerph14111355)] [Medline: [29112126](https://pubmed.ncbi.nlm.nih.gov/29112126/)]
17. Frank TD, Carter A, Jahagirdar D. Global, regional, and national incidence, prevalence, and mortality of HIV, 1980-2017, and forecasts to 2030, for 195 countries and territories: a systematic analysis for the global burden of diseases, injuries, and risk factors study 2017. *Lancet HIV*. Dec 2019;6(12):e831-e859. [doi: [10.1016/S2352-3018\(19\)30196-1](https://doi.org/10.1016/S2352-3018(19)30196-1)] [Medline: [31439534](https://pubmed.ncbi.nlm.nih.gov/31439534/)]
18. Mounier-Jack S, Mayhew SH, Mays N. Integrated care: learning between high-income, and low- and middle-income country health systems. *Health Policy Plan*. Nov 1, 2017;32(suppl\_4):iv6-iv12. [doi: [10.1093/heapol/czx039](https://doi.org/10.1093/heapol/czx039)] [Medline: [29194541](https://pubmed.ncbi.nlm.nih.gov/29194541/)]
19. Pincombe M, Reese V, Dolan CB. The effectiveness of national-level containment and closure policies across income levels during the COVID-19 pandemic: an analysis of 113 countries. *Health Policy Plan*. Aug 12, 2021;36(7):1152-1162. [doi: [10.1093/heapol/czab054](https://doi.org/10.1093/heapol/czab054)] [Medline: [33942081](https://pubmed.ncbi.nlm.nih.gov/33942081/)]
20. Cooper V, Clatworthy J, Youssef E, et al. Which aspects of health care are most valued by people living with HIV in high-income countries? a systematic review. *BMC Health Serv Res*. Nov 30, 2016;16(1):677. [doi: [10.1186/s12913-016-1914-4](https://doi.org/10.1186/s12913-016-1914-4)] [Medline: [27899106](https://pubmed.ncbi.nlm.nih.gov/27899106/)]
21. Park S, Park J. Identifying the knowledge structure and trends of outreach in public health care: a text network analysis and topic modeling. *Int J Environ Res Public Health*. Sep 3, 2021;18(17):9309. [doi: [10.3390/ijerph18179309](https://doi.org/10.3390/ijerph18179309)] [Medline: [34501897](https://pubmed.ncbi.nlm.nih.gov/34501897/)]

22. Subakti A, Murfi H, Hariadi N. The performance of BERT as data representation of text clustering. *J Big Data*. 2022;9(1):15. [doi: [10.1186/s40537-022-00564-9](https://doi.org/10.1186/s40537-022-00564-9)] [Medline: [35194542](https://pubmed.ncbi.nlm.nih.gov/35194542/)]
23. Thakur K, Kumar V. Application of text mining techniques on scholarly research articles: methods and tools. *New Rev Acad Librariansh*. Jul 3, 2022;28(3):279-302. [doi: [10.1080/13614533.2021.1918190](https://doi.org/10.1080/13614533.2021.1918190)]
24. Dancy-Scott N, Dutcher GA, Keselman A, et al. Trends in HIV terminology: text mining and data visualization assessment of international AIDS conference abstracts over 25 years. *JMIR Public Health Surveill*. May 4, 2018;4(2):e50. [doi: [10.2196/publichealth.8552](https://doi.org/10.2196/publichealth.8552)] [Medline: [29728344](https://pubmed.ncbi.nlm.nih.gov/29728344/)]
25. Devlin J, Chang MW, Lee K, Toutanova K. Bert: pretraining of deep bidirectional transformers for language understanding. *arXiv*. Preprint posted online on 2018. [doi: [10.48550/arXiv.1810.04805](https://doi.org/10.48550/arXiv.1810.04805)]
26. Xu WW, Tshimula JM, Dubé È, et al. Unmasking the Twitter discourses on masks during the COVID-19 pandemic: user cluster-based BERT topic modeling approach. *JMIR Infodemiology*. 2022;2(2):e41198. [doi: [10.2196/41198](https://doi.org/10.2196/41198)] [Medline: [36536763](https://pubmed.ncbi.nlm.nih.gov/36536763/)]
27. El-Gayar O, Al-Ramahi M, Wahbeh A, Nasrallah T, Elnoshokaty A. A comparative analysis of the interpretability of LDA and LLM for topic modeling: the case of healthcare apps. Presented at: AMCIS 2024 Proceedings; Aug 15-17, 2024; Salt Lake City, Utah. 2024. URL: [https://aisel.aisnet.org/amcis2024/health\\_it/health\\_it/22](https://aisel.aisnet.org/amcis2024/health_it/health_it/22) [Accessed 2025-06-19]
28. Egger R, Yu J. A topic modeling comparison between LDA, NMF, Top2Vec, and BERTopic to demystify Twitter posts. *Front Sociol*. 2022;7:886498. [doi: [10.3389/fsoc.2022.886498](https://doi.org/10.3389/fsoc.2022.886498)] [Medline: [35602001](https://pubmed.ncbi.nlm.nih.gov/35602001/)]
29. Raman R, Pattnaik D, Hughes L, Nedungadi P. Unveiling the dynamics of AI applications: a review of reviews using scientometrics and BERTopic modeling. *J Innov Knowl*. Jul 2024;9(3):100517. [doi: [10.1016/j.jik.2024.100517](https://doi.org/10.1016/j.jik.2024.100517)]
30. Baird A, Xia Y, Cheng Y. Consumer perceptions of telehealth for mental health or substance abuse: a Twitter-based topic modeling analysis. *JAMIA Open*. Jul 2022;5(2):ooac028. [doi: [10.1093/jamiaopen/ooac028](https://doi.org/10.1093/jamiaopen/ooac028)] [Medline: [35495736](https://pubmed.ncbi.nlm.nih.gov/35495736/)]
31. Chandra R, Saini R. Biden vs Trump: modeling US general elections using BERT language model. *IEEE Access*. 2021;9:128494-128505. [doi: [10.1109/ACCESS.2021.3111035](https://doi.org/10.1109/ACCESS.2021.3111035)]
32. Karabacak M, Jagtiani P, Jain A, Panov F, Margetis K. Tracing topics and trends in drug-resistant epilepsy research using a natural language processing-based topic modeling approach. *Epilepsia*. Apr 2024;65(4):861-872. [doi: [10.1111/epi.17890](https://doi.org/10.1111/epi.17890)] [Medline: [38314969](https://pubmed.ncbi.nlm.nih.gov/38314969/)]
33. Caskey J, McConnell IL, Oguss M, et al. Identifying COVID-19 outbreaks from contact-tracing interview forms for public health departments: development of a natural language processing pipeline. *JMIR Public Health Surveill*. Mar 8, 2022;8(3):e36119. [doi: [10.2196/36119](https://doi.org/10.2196/36119)] [Medline: [35144241](https://pubmed.ncbi.nlm.nih.gov/35144241/)]
34. WHO. Consolidated guidelines on the use of antiretroviral drugs for treating and preventing HIV infection: recommendations for a public health approach. 2013. URL: <https://www.who.int/publications/i/item/9789241505727> [Accessed 2025-06-20]
35. Asmussen CB, Møller C. Smart literature review: a practical topic modelling approach to exploratory literature review. *J Big Data*. Dec 2019;6(1):1-18. [doi: [10.1186/s40537-019-0255-7](https://doi.org/10.1186/s40537-019-0255-7)]
36. Karabacak M, Schupper AJ, Carr MT, Hickman ZL, Margetis K. From text to insight: a natural language processing-based analysis of topics and trends in neurosurgery. *Neurosurgery*. Apr 1, 2024;94(4):679-689. [doi: [10.1227/neu.000000000002763](https://doi.org/10.1227/neu.000000000002763)] [Medline: [37988054](https://pubmed.ncbi.nlm.nih.gov/37988054/)]
37. Wilianto D, Suganda Girsang A. Automatic short answer grading on high school's e-learning using semantic similarity methods. *TEM Journal*. Feb 27, 2023;12(1):297-302. [doi: [10.18421/TEM121-37](https://doi.org/10.18421/TEM121-37)]
38. Hatvani P, Yang ZG. Training embedding models for Hungarian. Presented at: 2024 IEEE 3rd Conference on Information Technology and Data Science (CITDS); Aug 26-28, 2024; Debrecen, Hungary. [doi: [10.1109/CITDS62610.2024.10791389](https://doi.org/10.1109/CITDS62610.2024.10791389)]
39. Hidayat AA, Nirwantono R, Budiarto A, Pardamean B. BERT-based topic modeling approach for malaria research publication. Presented at: 2022 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS); Nov 16-17, 2022; Jakarta, Indonesia. [doi: [10.1109/ICIMCIS56303.2022.10017743](https://doi.org/10.1109/ICIMCIS56303.2022.10017743)]
40. Sangaraju VR, Bolla BK, Nayak DK. KhJ topic modelling on consumer financial protection bureau data: an approach using BERT based embeddings. *arXiv*. Preprint posted online on 2022. [doi: [10.48550/arXiv.2205.07259](https://doi.org/10.48550/arXiv.2205.07259)]
41. McInnes L, Healy J, Melville J. UMAP: uniform manifold approximation and projection for dimension reduction. *arXiv*. 2018. [doi: [10.48550/arXiv.1802.03426](https://doi.org/10.48550/arXiv.1802.03426)]
42. McInnes L, Healy J, Astels S. HdbSCAN: Hierarchical density based clustering. *J Open Source Softw*. 2017;2(11):205. [doi: [10.21105/joss.00205](https://doi.org/10.21105/joss.00205)]
43. Grootendorst M. BERTopic: neural topic modeling with a class-based TF-IDF procedure. *arXiv*. Preprint posted online on 2022. [doi: [10.48550/arXiv.2203.05794](https://doi.org/10.48550/arXiv.2203.05794)]
44. Zhao W, Chen JJ, Perkins R, et al. A heuristic approach to determine an appropriate number of topics in topic modeling. *BMC Bioinformatics*. 2015;16(Suppl 13):S8. [doi: [10.1186/1471-2105-16-S13-S8](https://doi.org/10.1186/1471-2105-16-S13-S8)] [Medline: [26424364](https://pubmed.ncbi.nlm.nih.gov/26424364/)]

45. Riaz A, Abdulkader O, Ikram MJ, Jan S. Exploring topic modelling: a comparative analysis of traditional and transformer-based approaches with emphasis on coherence and diversity. *Int J Electrical Computer Engineer (IJECE)*. 2025;15(2):1933. [doi: [10.11591/ijece.v15i2.pp1933-1948](https://doi.org/10.11591/ijece.v15i2.pp1933-1948)]
46. Li S, Pan R, Luo H, Liu X, Zhao G. Adaptive cross-contextual word embedding for word polysemy with unsupervised topic modeling. *Knowl Based Syst*. Apr 2021;218:106827. [doi: [10.1016/j.knosys.2021.106827](https://doi.org/10.1016/j.knosys.2021.106827)]
47. Murthy D, Keshari S, Arora S, et al. Categorizing E-cigarette-related tweets using BERT topic modeling. *Emerg Trends Drugs Addict Health*. Dec 2024;4:100160. [doi: [10.1016/j.etched.2024.100160](https://doi.org/10.1016/j.etched.2024.100160)]
48. Adam R, Kogler ML. Bidirectional topic matching: quantifying thematic overlap between corpora through topic modelling. *arXiv*. Preprint posted online on 2024. [doi: [10.48550/arXiv.2412.18376](https://doi.org/10.48550/arXiv.2412.18376)]
49. WHO. Countries. URL: <https://www.who.int/countries> [Accessed 2025-06-20]
50. Allegrante JP, Wells MT, Peterson JC. Interventions to support behavioral self-management of chronic diseases. *Annu Rev Public Health*. Apr 1, 2019;40(1):127-146. [doi: [10.1146/annurev-publhealth-040218-044008](https://doi.org/10.1146/annurev-publhealth-040218-044008)] [Medline: [30601717](https://pubmed.ncbi.nlm.nih.gov/30601717/)]
51. Kay ES, Batey DS, Mugavero MJ. The HIV treatment cascade and care continuum: updates, goals, and recommendations for the future. *AIDS Res Ther*. Dec 2016;13(1):1-7. [doi: [10.1186/s12981-016-0120-0](https://doi.org/10.1186/s12981-016-0120-0)]
52. Zeluf-Andersson G, Eriksson LE, Schönnesson LN, Höijer J, Måneshall P, Ekström AM. Beyond viral suppression: the quality of life of people living with HIV in Sweden. *AIDS Care*. Apr 2019;31(4):403-412. [doi: [10.1080/09540121.2018.1545990](https://doi.org/10.1080/09540121.2018.1545990)] [Medline: [30444140](https://pubmed.ncbi.nlm.nih.gov/30444140/)]
53. Lazarus JV, Wohl DA, Cascio M, et al. Long-term success for people living with HIV: a framework to guide practice. *HIV Med*. Mar 2023;24(supplement 2):8-19. [doi: [10.1111/hiv.13460](https://doi.org/10.1111/hiv.13460)] [Medline: [36920412](https://pubmed.ncbi.nlm.nih.gov/36920412/)]
54. Guaraldi G, Arends J, Buhk T, et al. “Moving fourth”: a vision toward achieving healthy living with HIV beyond viral suppression. *AIDS Rev*. 2019;21(3):135-142. [doi: [10.24875/AIDSRev.19000088](https://doi.org/10.24875/AIDSRev.19000088)] [Medline: [31532396](https://pubmed.ncbi.nlm.nih.gov/31532396/)]
55. Centers for Disease Control and Prevention. Estimated HIV incidence and prevalence in the united states. 2017–2021. *HIV Surveillance Supplemental Report*; 2023. URL: <https://stacks.cdc.gov/view/cdc/149080> [Accessed 2025-06-23]
56. Du M, Liu M, Liu J. Global, regional, and national disease burden and attributable risk factors of HIV/AIDS in older adults aged 70 years and above: a trend analysis based on the Global Burden of Disease study 2019. *Epidemiol Infect*. Dec 15, 2023;152:e2. [doi: [10.1017/S0950268823001954](https://doi.org/10.1017/S0950268823001954)] [Medline: [38097398](https://pubmed.ncbi.nlm.nih.gov/38097398/)]
57. Bor J, Fischer C, Modi M, et al. Changing knowledge and attitudes towards HIV treatment-as-prevention and “undetectable = untransmittable”: a systematic review. *AIDS Behav*. Dec 2021;25(12):4209-4224. [doi: [10.1007/s10461-021-03296-8](https://doi.org/10.1007/s10461-021-03296-8)] [Medline: [34036459](https://pubmed.ncbi.nlm.nih.gov/34036459/)]
58. Hussen SA, Chahroudi A, Boylan A, Camacho-Gonzalez AF, Hackett S, Chakraborty R. Transition of youth living with HIV from pediatric to adult-oriented healthcare: a review of the literature. *Future Virol*. Oct 2014;9(10):921-929. [doi: [10.2217/fvl.14.73](https://doi.org/10.2217/fvl.14.73)]
59. Lazarus JV, Safreed-Harmon K, Kamarulzaman A, et al. Consensus statement on the role of health systems in advancing the long-term well-being of people living with HIV. *Nat Commun*. 2021;12(1):4450. [doi: [10.1038/s41467-021-24673-w](https://doi.org/10.1038/s41467-021-24673-w)]
60. Duthely LM, Sanchez-Covarrubias AP. Digitized HIV/AIDS treatment adherence interventions: a review of recent SMS/texting mobile health applications and implications for theory and practice. *Front Commun (Lausanne)*. Nov 2020;5:530164. [doi: [10.3389/fcomm.2020.530164](https://doi.org/10.3389/fcomm.2020.530164)] [Medline: [33644162](https://pubmed.ncbi.nlm.nih.gov/33644162/)]
61. Simoni JM, Kutner BA, Horvath KJ. Opportunities and challenges of digital technology for HIV treatment and prevention. *Curr HIV/AIDS Rep*. Dec 2015;12(4):437-440. [doi: [10.1007/s11904-015-0289-1](https://doi.org/10.1007/s11904-015-0289-1)] [Medline: [26412082](https://pubmed.ncbi.nlm.nih.gov/26412082/)]
62. Haris M, Abbas R. Four decades of HIV: global trends, testing assays, treatment, and challenges. *Zoonoses*. 2024;4(1):997. [doi: [10.15212/ZOONOSES-2023-0039](https://doi.org/10.15212/ZOONOSES-2023-0039)]
63. McCree DH, Byrd KK, Johnston M, Gaines M, Weidle PJ. Roles for pharmacists in the “ending the HIV epidemic: a plan for America” initiative. *Public Health Rep*. 2020;135(5):547-554. [doi: [10.1177/0033354920941184](https://doi.org/10.1177/0033354920941184)] [Medline: [32780671](https://pubmed.ncbi.nlm.nih.gov/32780671/)]
64. Crawford ND, Myers S, Young H, Klepser D, Tung E. The role of pharmacies in the HIV prevention and care continuums: a systematic review. *AIDS Behav*. Jun 2021;25(6):1819-1828. [doi: [10.1007/s10461-020-03111-w](https://doi.org/10.1007/s10461-020-03111-w)] [Medline: [33386509](https://pubmed.ncbi.nlm.nih.gov/33386509/)]
65. Brennan-Ing M, Mattas E. Aging with HIV: working to ensure equity and inclusion. *Gerontology*. 2023;69(8):946-952. [doi: [10.1159/000530723](https://doi.org/10.1159/000530723)] [Medline: [37071979](https://pubmed.ncbi.nlm.nih.gov/37071979/)]
66. Tam G, Wong NS, Lee SS. A survey of the involvement of primary care doctors in HIV prevention and care in a low-prevalence, high-income setting. *BMC Fam Pract*. Jan 28, 2021;22(1):27. [doi: [10.1186/s12875-021-01376-1](https://doi.org/10.1186/s12875-021-01376-1)] [Medline: [33504326](https://pubmed.ncbi.nlm.nih.gov/33504326/)]

67. Bond KT, Chandler R, Chapman-Lambert C, et al. Applying a nursing perspective to address the challenges experienced by cisgender women in the HIV status neutral care continuum: a review of the literature. *J Assoc Nurses AIDS Care*. 2021;32(3):283-305. [doi: [10.1097/JNC.000000000000243](https://doi.org/10.1097/JNC.000000000000243)] [Medline: [33929979](https://pubmed.ncbi.nlm.nih.gov/33929979/)]
68. WHO. Public health milestones through the years. 2024. URL: <https://www.who.int/campaigns/75-years-of-improving-public-health/milestones#year-1945> [Accessed 2025-06-20]

---

## Abbreviations

**AMR:** Region of the Americas  
**ART:** antiretroviral therapy  
**BERT:** bidirectional encoder representations from transformers  
**LDA:** latent dirichlet allocation  
**QoL:** quality of life  
**TF-IDF:** term frequency-inverse document frequency  
**WHO:** World Health Organization

---

*Edited by Tsz Ho Kwan; peer-reviewed by Don C D E S Jarlais, Rumana Saifi; submitted 04.08.2024; final revised version received 07.04.2025; accepted 28.04.2025; published 09.07.2025*

*Please cite as:*

*Lee S, Kim L, Shim MS, Kim GS*

*Identifying Health Care Services Offered in the HIV Care Continuum via a Machine Learning–Based Topic Modeling Approach: Exploratory Literature Review*

*JMIR Public Health Surveill 2025;11:e65081*

*URL: <https://publichealth.jmir.org/2025/1/e65081>*

*doi: [10.2196/65081](https://doi.org/10.2196/65081)*

© SangA Lee, Layoung Kim, Mi-So Shim, Gwang Suk Kim. Originally published in *JMIR Public Health and Surveillance* (<https://publichealth.jmir.org>), 09.07.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in *JMIR Public Health and Surveillance*, is properly cited. The complete bibliographic information, a link to the original publication on <https://publichealth.jmir.org>, as well as this copyright and license information must be included.