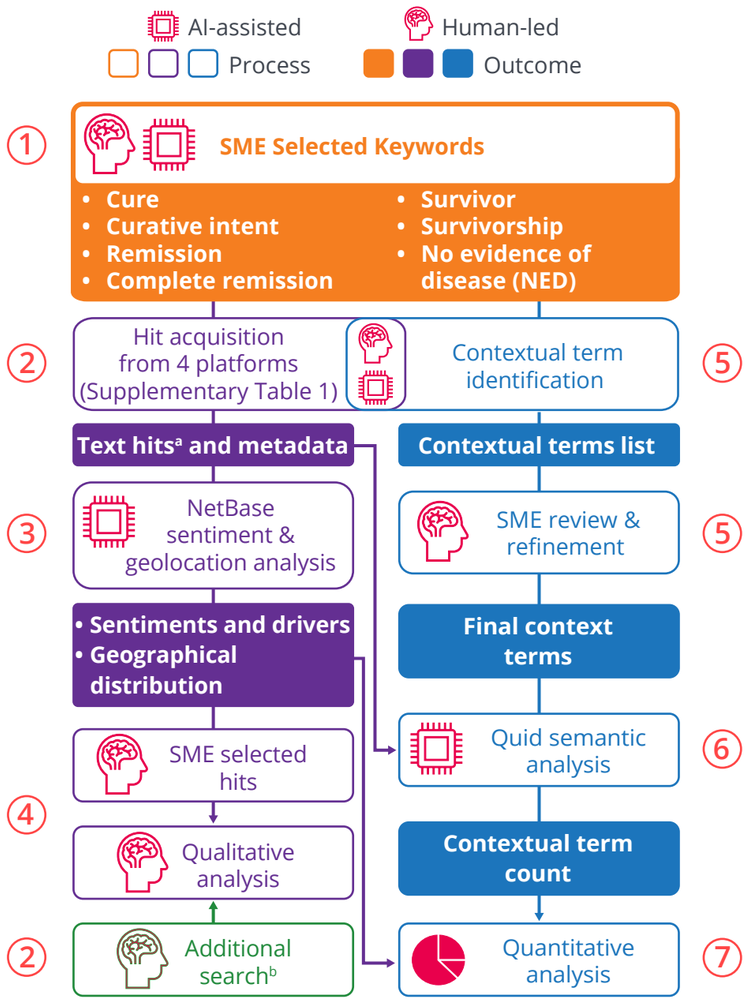
Artificial Intelligence in Performing Landscape Review and Linguistic Analysis for Curative Intent in Prostate Cancer

Supplementary Material

SUPPLEMENTARY FIGURE 1



AI, artificial intelligence.
[a]May have >1 keyword and/or disease area.
[b]Manual search for clinical guidelines and health technology assessments.

Step	Process Step Description
1	SMEs developed a list of keywords related to cure using Elicit
2	Hits were acquired for each keyword searched in 4 stakeholder platforms by SMEs (academic researchers, healthcare professionals [HCPs], and policymakers) and by NetBase (the general public) If gaps in policymaker documents were noted, additional targeted, SME-led searches were conducted. Keyword searches were done in conjunction with terms related to prostate cancer
3	NetBase analyzed hits related to each keyword to identify the overall sentiment and drivers (attributes, emotions, behaviors, things) contributing to the sentiment NetBase analyzed geographical data from social media posts
4	SMEs reviewed the NetBase "things" subset of sentiment drivers to identify those of interest. NetBase then extracted related hits with ≥1 sentiment driver. SMEs analyzed the hits for themes and concepts associated with the keywords SMEs also analyzed additional policymaker documents acquired from manual searches from step 2
5	SMEs asked the Elicit program key questions such as "What is the definition of <*keyword*> in prostate cancer?" to shape the list of contextual terms and refined this to a final list
6	Quid performed an automated semantic analysis of unstructured text and identified clusters of hits based on similar semantics and word use
7	Quantitative analysis of sentiments and drivers was performed using NetBase Contextual term count was performed using a custom R script

Artificial Intelligence in Performing Landscape Review and Linguistic Analysis for Curative Intent in Prostate Cancer

Supplementary Material

SUPPLEMENTARY TABLE 1: Platforms used for keyword search

Platform (stakeholder) and document types	Timeframe	Dates
MEDLINE (academic researchers)		
Published, peer-reviewed literature	5 years	From 2017 to 30 March 2023
Sermo (HCPs)		
Closed discussion forum for registered HCPs	2 years[a]	From 23 March 2021 to 23 March 2023
Overton (policymakers)		
Policy documents (eg, healthcare technology assessments, guidelines, etc)	5 years[a]	From 31 March 2018 to 31 March 2023
Social media[b] (general public)		
Twitter[c], Reddit, blogs etc, by the general public (patients, caregivers, HCPs, and patient advocates)	27 months[a]	From 20 January 2021 to 20 April 2023

[a]Maximum available timeframe at the time of the study.
[b]Social media names are as of at the time of the study.
[c]Twitter (now X) utilizes a Decahose scaling method that provides 10% of random tweets associated with the keywords.

SUPPLEMENTARY TABLE 2: Keywords and contextual terms

Keywords	Contextual terms
• Cure • Survivor • Remission • Complete remission • Survivorship • Curative intent • NED	• PSA/Prostate-specific antigen • Gleason • Cancerous cell(s) • Surgical/Surgery/Prostatectomy • Disease manifestation(s) • Biochemical • Palpable • Rectal exam(s) • Nonmetastatic/Non-metastatic • Resectable • Expectant • Indolent • Localized/Localised • Locally advanced/Locally-advanced • nmCRPC • nmCSPC • mCRPC • mCSPC • CRPC • CSPC • Stage III/Stage 3 • Stage IIIA/Stage 3A • Stage IIIB/Stage 3B • Stage IIIC/Stage 3C • 4+3 • 3+4 • 4+4 • 4 3 • 3 4 • 4 4

Artificial Intelligence in Performing Landscape Review and Linguistic Analysis for Curative Intent in Prostate Cancer

Supplementary Material

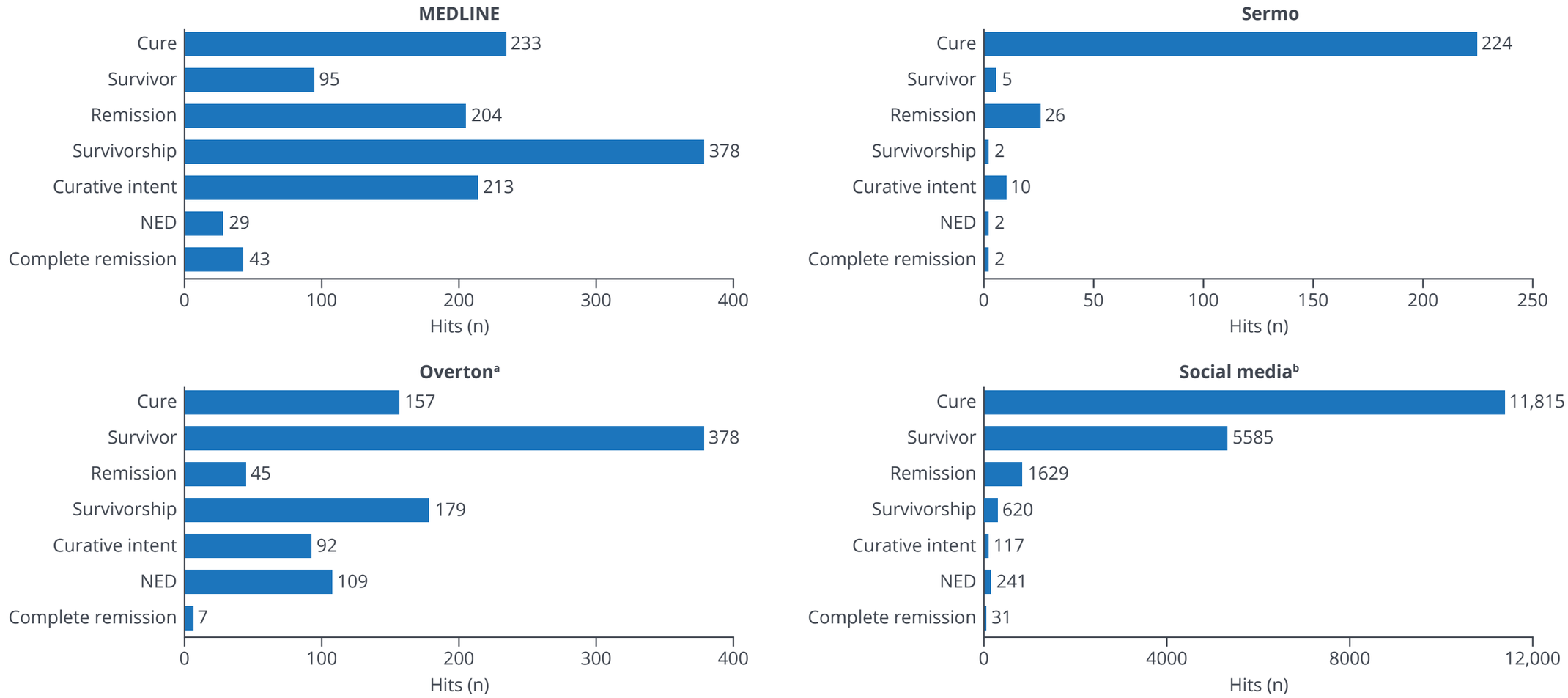
SUPPLEMENTARY TABLE 3: Glossary

Term	Context
Elicit	A semantic search engine utilizing a large language model that allows natural-language queries based on the semantic nature of the question
Contextual term	Terms used frequently with the keywords
Hit	Text from abstracts, documents, posts, replies, and comments related to selected keywords
NetBase Quid	A proprietary fully integrated AI engine that uses natural language processing for linguistic analysis of structured and unstructured data, including social media. NetBase Quid consists of 2 platforms, the social media analytics platform (NetBase) and the AI-driven text analytic platform (Quid)
Sentiments	The NetBase output reflecting the feelings associated with each keyword. Sentiments can be positive or negative and are extracted from drivers
Drivers	The NetBase output consisting of top trending terms associated with each keyword. Drivers are classified into 4 categories: attributes, emotions, behaviors, and things
Attributes	Positives or negatives, such as likes or dislikes
Emotions	Positive or negative feelings, such as “love” or “hate”
Behaviors	Positive or negative actions, such as “buy” or “avoid”
Things	Frequently occurring objects of sentiment (location, brand, people)
Decahose	A scaling method utilized by Twitter (now X) that provides 10% of random tweets associated with the keywords

Artificial Intelligence in Performing Landscape Review and Linguistic Analysis for Curative Intent in Prostate Cancer

Supplementary Material

SUPPLEMENTARY FIGURE 2



[a]Multiple hits could be returned from a single policy document.
[b]Number of hits are approximate due to NetBase Twitter (now X) decahose scaling.