

厚生労働科学研究費補助金（がん政策研究事業）
分担研究報告書

一般向け各種がん情報の構造化の試み

研究協力者 木下 乙女 国立がん研究センターがん対策情報センターがん情報提供部
研究協力者 早川 雅代 国立がん研究センターがん対策情報センターがん情報提供部
代表研究者 高山 智子 国立がん研究センターがん対策情報センターがん情報提供部

研究要旨

がん患者や一般市民向けに最新の信頼できる情報を持続的に提供するためには『質の高い情報を、効率良く継続的に更新する手法』の検討が必要である。本研究では、一般向けがん情報の主軸である各種がんに関する情報を構造化することにより、その手法を模索することを目的とした。

既存情報を基に検討し、構成要素を検討した。作成した構造化フォーマットにより情報を試作した。質を担保しつつ、「短期間で作成可能とする」、「情報媒体別の情報を一元的に作成する」、「厳選した文字数で作成する」、「経験の浅い作成者でも作成可能とする」ことを達成できる可能性が示唆された。また、最小限の文字数で作成することにより情報の可読性や検索性も向上したと考えられ、情報を構造化することのさらなる意義が示唆された。

A. 研究目的

近年の医療技術の進歩等により、がん患者をはじめとした一般市民向けのがん情報の提供においては、求められる情報量が急増している。必然的に必要とされる更新頻度も増加している。また、情報量の増加による情報の可読性や検索性の低下などの問題点も指摘されつつある。さらに、可読性の面では、近年では、スマートフォンでの情報利用者がパソコンでの利用者を上回っていることから¹⁾、最小限の文字数で質の高い情報を提供することが求められつつある。

これらの課題を解決し、一般向けに最新の信頼できる情報を持続的に提供するため

には、情報作成の担い手を増やすこととともに、効率良く、厳選した情報を質の高い状態で更新する手法の検討が必要である。

本研究では、『質の高い情報を、効率良く継続的に更新する手法』を検討するために、一般向けがん情報の主軸である各種がんに関する情報の構造化を試みた。構造化により、情報の質を担保しつつ、効率的に情報を作成するために、「短期間で作成可能とする」「情報媒体別の情報を一元的に作成する」「厳選した文字数で作成する」ことを目的とした。また、継続的に更新するための情報作成の担い手を増やすために「経験の浅い作成者でも作成可能とする」ことも目的とした。

B. 研究方法

本研究では、構造化の検討用素材として、国立がん研究センターがん対策情報センターが運営するがん情報サービスの各種がんに関する情報を用いた。

なお、本研究は人を対象にしないという理由から、研究倫理審査の申請は不要と判断した。

(1) 全体文字数の検討

2017年4月時点で、ウェブサイトに掲載されていた情報の中で、2013年に作成された全6種（喉頭がん、中下咽頭がん、上咽頭がん、中皮腫、胸腺腫と胸腺がん、精巣腫瘍）及び2015年～2017年に作成された部位や罹患率などが異なる8種（乳がん、大腸がん、胃がん、肝細胞がん、膀胱がん、軟部肉腫、多発性骨髄腫、小児白血病）及びの全体文字数を調査した。がん情報サービス運営者に、これまでの全体文字数に関する外部からの声をヒアリングした。

(2) 情報の構成要素の検討

情報の構成要素を抽出するため、3種（乳がん、大腸がん、多発性骨髄腫）のウェブサイトの情報より段落ごとに文章を切り出し、この段落を情報の構成要素の最小単位（1ピース）とし、それぞれ要素名を付与し整理した。

各ピースの必要性について、検討チーム（一般向け各種がんに関する情報を作成した経験のある5名〔薬剤師1名、看護師2名、臨床心理士1名、化学修士1名〕）による議論により、表1の4種類に各要素を分類した。

表1 情報の構成要素の分類

①	共通ピース	原則全がん種で必要と考えられるピース
②	がん種特有ピース	がん種によって必要となるピース
③	削除または統合ピース	共通もしくはがん種特有ピースに統合可能なピース または 不要なピース
④	追加ピース	上記ピースには含まれないが追加すべきピース

(3) 構成要素の各文字数の検討と全体文字数の調整

3種の情報の1ピースあたりの文字数を調査し、スマートフォンでの使用を考慮して、目安とする文字数を検討した。

(2)で分類した項目と1ピースあたりの文字数により、全体文字数を再検討した。

(4) 紙媒体用構成要素の検討

ウェブサイトに対して紙媒体である小冊子では文字数が制限されることから、情報量を減らす必要が生じる。情報作成の効率化のためには、異なる情報媒体でも同一の情報を使用することが望ましい。そのため冊子として必要な構成要素を表2による分類を検討した。

表2 小冊子に使用する構成要素の分類

⑤	冊子共通ピース	冊子に必須とするピース
⑥	冊子がん種特有ピース	がん種によって必要であり、冊子にも必要であるピース
⑦	冊子不使用ピース	冊子には掲載しないピース

(5) 構造化フォーマットの作成

(1)～(4)の検討結果に基づいて構造化フォーマット(エクセル版)を作成した。また、各ピースのうち、記述を統一できるピース(統計、生存率など)については、がん種横断的に情報の欠落が生じないこと及びユーザーの混乱を最小限に抑えることを目的として統一文書を組み込んだ。

(6) 構造化フォーマットの試用

作成した構造化フォーマットを用いて、「前立腺がん」「膵臓がん」「外陰がん」の情報作成を行い、専門家及び患者等に査読を依頼した。作成期間、査読による内容の確認により、構造化が情報の質を担保しつつ、課題の解決につながるかについて検討した。

さらに、作成した情報をがん情報サービスの掲載情報と比較した。

C. 研究結果

(1) 全体文字数の検討

2013年に作成された6種の情報の平均全文字数は、11,713字、最低10,364字、最大13,907字であった。一方で2015年～2017年に作成された8種の情報の平均全文字数は、20,177字、最低9,629字、最大29,706字であり、がん種の違いを考慮しても、近年の情報量の増加により新しく作成された情報は文字数が明らかに多い傾向にあった。また、がん情報サービス運営者へのヒアリングでは、近年1つのがん種で2万字を超える情報が増えたことにより、査読者等

から文字数が多すぎるとの指摘が多かったとの報告があった。

上記の結果より、全体文字数としては、2,400字程度でがん種横断的に記載されていた「診療の流れ」などの情報は不要であることを考慮して、約10,000字を目指すことが妥当と考えられた。

また、詳細検討に使用する3種のがん情報の文字数は、乳がん29,555字、大腸がん23,686字、多発性骨髄腫18,950字であった。

(2) 情報の構成要素の検討

各情報を段落ごとに抽出した結果、ピース数は乳がん103、大腸がん114、多発性骨髄腫95であった。類似した要素をまとめた結果、218ピースが全要素として抽出された。

抽出されたピースについて、4段階に分類した結果の概要を表3に示した。また、全結果を附表1に示した。①共通42ピース、②がん種特有ピース64ピースであった。③削除に分類したものは優先度が低いと考えられた「罹患率の国際比較」などの8ピースとがん情報全般に関する4ピース(診療の流れ、受診と相談の勧め、がんと言われたとき、自分にあった治療法を考える)であった。

また、上記3種の構成要素としては含まれていない④追加ピースとして、近年情報が充実しつつあり、患者への情報提供の必要性が高まっていた「リハビリテーション」を要素として追加した。

表3 構成要素の分類結果

分類	ピース数	内訳	項目
共通ピース	42		臓器の構造、罹患率、生存率など
がん種特有ピース	64		病型分類、検診の内容、内視鏡治療など
削除または統合ピース	112	必須・任意項目に集約・統合(100)	緩和ケア 4 ピースを「緩和ケア」(必須 1 ピース)に集約、日和見感染・食事・運動など 25 ピースを「日常生活をおくる上で」に統合(必須 2 ピース、任意 6 ピース)、定期検査 8 ピースを「経過観察」(必須 2 ピース)に集約、など
		本コンテンツには不要(12)	<削除したもの；8 ピース>がんの全体的な経過、罹患率の国際比較、死亡率、社会復帰(社会復帰を考える時期、体調が優れないときの相談等 5 ピース) <別ページ「治療にあたって」を作成しリンクしたもの；4 ピース>診療の流れ、受診と相談の勧め、がんと言われたとき、自分に合った治療法を考える
追加ピース	4		リハビリテーション
合計	222		

(3) 構成要素の各文字数の検討と全体文字数の調整

3 種の 1 ピースあたりの文字数の平均値は 177 字であった。この文字数は、一般的な画面サイズのスマートフォン画面で、1 画面で表示される 7 行程度の文字数であったことから、1 ピースあたりの目安の文字数を 150 字と定めた。

この文字数は、PC 画面で 3 行程度で表示され、検討チーム全員が可読性が良いと判断した。

(2) で分類した①共通ピース、②がん種特有ピース、④追加ピースの合計ピース数を全て使用した場合の全体文字数は 16,500 字となり、(1)の結果の文字数(約 10,000 字)

を大幅に超過していた。全体文字数を減らすために、150 字の説明が不要な要素について、1 要素 2/3 ピース(100 字)の項目を設定した。なお、がん種特有ピースは、全ピースを使用することはないことから、全体文字数が 10,000 字に近づくよう 30 ピースを目安とした。よって、全体文字数は、共通 42 ピース分の文字数(6300) + がん種特有 30 ピース分の文字数(4500) = 10800 字とした。

(4) 紙媒体用構成要素の検討

文字数制限のある紙媒体である小冊子を作成する際に抽出するピース数を検討した。構成要素を減らすために、対象をがんと診断されたばかりの患者を想定して分類した。

④冊子共通 30 ピース、⑤冊子がん種特有 38 ピース、⑥冊子では不使用 42 ピースとなった(全 110 ピース)。共通のピースでも冊子共通ピースとしなかったピースがあった(例：生存率、緩和ケア、臨床試験など)。

(5) 構造化フォーマットの作成

作成した構造化フォーマットの一部を図 1 に示す。構造化フォーマットは、構成要素や文字数を確認しながら情報を入力する様式とした。

また、統一文書を 12 種(罹患率、予防、検診、妊娠や出産について、生存率、緩和ケア、臨床試験、転移・再発とは、など)作成し、あらかじめ構造化フォーマットに組み込んだ(図 1 青地部)。

図 1 構造化フォーマット(一部抜粋)

番号	見出し	項目	ピース数	WEB 必須の件数	冊子必須の件数	目次の文字数	旧版数	文字数(旧)	新規構築	文字数(新)	参考文数
D05	薬物療法	薬剤の種類、副作用、その他	1	△	△	150		0		0	
D00	臨床試験	●●●ががんの臨床試験	1	◎	×	150		0	現在国内で行われている臨床試験(治験)研究が実施する癌種は前立腺、および乳癌企業や医師が実施する治験の一部に限って、「がんの臨床試験」で情報を構築することができます。	143	
D01	転移・再発とは	転移・再発とは	1	◎	◎	100		0	転移とは、がん細胞がリンパ液や血液の流れなどに乗って別の臓器に移動し、そこで増殖することです。また、再発とは、治療の効果がよりがんがなくなったあと、再びがんが出現することです。	100	
D02	転移・再発とは	転移・再発とは	1	◎	◎	200		0		0	

(6) 構造化フォーマットの試用

作成した構造化フォーマットを用いて、「前立腺がん」「膵臓がん」「外陰がん」を作成した結果、3 種とも約 3 ヶ月で情報の作成が可能であった。がん情報サービス運営者より報告があったがん情報サービスでの平均作成期間 6 カ月と比較して約 3 ヶ月

程度短縮した。

文字数について、表 4 に示す。希少がんであったことから文字数が少なかった「外陰がん」は増加し、文字数が多かった「前立腺がん」「膵臓がん」は減少した。

表 4 本研究で作成したがん種の文字数の比較

	改訂前全体文字数	改訂後全体文字数	差(増減割合)
前立腺がん	16860	12444	-4416 (-26%)
膵臓がん	13530	9942	-3588 (-27%)
外陰がん	3392	8554	+5162 (+152%)

ピース数の変化について、表 5 に示す。改訂前ピース数が少なかった「外陰がん」は増加し、多かった「膵臓がん」はやや減少した。「前立腺がん」は「リハビリテーション」の内容を多く組み入れたため、ピース数が増加した。がん種特有ピースは、全体文字数の 10,000 字に近づくよう 30 ピースを目安としていたが、前立腺がでは、41 ピースとなり、11 ピース超過した。一方で、他の 2 種では、30 ピースより少ないピース数となり、がん種特有ピースの目安を 30 ピースとしたことは妥当であることが示唆された。

表 5 本研究で作成したがん種のピース数の比較

	改訂前 ピース 数	改訂後ピース数				改訂前後の 合計ピース数の 差 (増減割合)
	合計	合計	共通 ピース	がん種特 有ピース	追加 ピース	
前立腺がん	62	86	42	41	3	+24 (+39%)
膵臓がん	70	66	42	23	1	-4 (-6%)
外陰がん	25	68	42	22	4	+43 (+172%)

専門家及び患者等による査読において、両媒体とも文字数が少ないことによる質の低下や情報不足に関する指摘は見られず、一定の質が担保されていると考えられた。

D. 考察

本検討では、各種がんに関する情報について、既存の情報を基に、構成要素としてがん種共通の要素とがん種特有の要素を抽出し、各文字数を規定することにより、情報の構造化を行った。実際に、構造化したフォーマットにより情報を作成することにより評価した結果、構造化は有用であることが示唆された。

構造化フォーマットを用いたことにより、情報の「短時間で作成可能とする」ことにつながり、情報を効率的に作成することができた。なお、情報の作成期間の短縮は、「情報媒体別の情報を一元的に作成した」によると考えられた。今後の更新では、薬物治療等、情報の変化が早い項目のピースのみを取り出して修正することにより、さらに情報更新の期間の短縮が可能になることが期

待される。

また、冊子作成においては、ピース間のピースを削除することから文章の流れが不自然になることが懸念されたが、構成要素単位で内容が完結することを目指して最初から文章を執筆することにつながり、かえって「厳選した文字数で作成する」ことにつながり、利用者の情報の可読性、検索性も向上したと考えられた。

さらに、本構造化では、共通項目により執筆することにより、がん種による情報の欠落を防ぐことができ、「経験の浅い作成者でも作成可能となる」と考えられた。

本検討により、文書の適切性の評価法 (SAM; Suitability Assessment of Materials) の評価項目の一部を満たし、読みやすい情報につながったと考える。しかしながら、満たしていない項目 (「文書の最後にまとめや要約があるか」、「情報が一方的に伝えられるのではなく、読み手が問題を解いたり質問に答えたりすることが求められているか」等) については、今後も検討の余地があると考えられた。

構造化での限界として、欠落情報のない質の高い情報作成につながる一方で、構造化に依存することによりがん種特有の必要な情報が欠落する可能性があることに注意が必要と考えられた。

海外の非営利団体が提供するがん情報サイトであるCancer Research UK でもWriter guidelines²⁾として同様の構成や項目立て（発生要因、症状、検査、病期、生存率など）を行っていた。また、本検討での1ピースに1つの要素を含めることは、「1段落に1つのポイントを述べる（We make one point per paragraph.）」点と一致していた。しかし、本検討では「1段落4文まで（We have a maximum of 4 sentences per paragraph.）」といった文の数に関する規定は設けておらず、今後、1段落に含む文の数についても検討の余地があると考えられた。

なお、本構造化フォーマットは、現在、本検討での3がん種以外にも、本研究班での取り組みの一つである全国がんセンター協議会での作成においても試用され、専門家からも構造化により情報が洗練されるとの意見があり、構造化の意義が見いだされた。

E. 結論

本研究では、がん患者や一般市民向けに最新の信頼できる情報を持続的に提供するために『質の高い情報を、効率良く継続的に更新する手法』の一つとして、一般向けの各種がんの情報の構造化を試みた。

既存情報を基に検討し、情報の最小単位として1ピース（150字）を設定し、構成要素を検討した。作成した構造化フォーマットにより情報を試作し、情報の質を担保

しつつ、「短期間で作成可能とする」、「情報媒体別の情報を一元的に作成する」、「厳選した文字数で作成する」、「経験の浅い作成者でも作成可能とする」ことが達成されることが示唆された。また、最小限の文字数で作成することにより情報の可読性や検索性も向上したと考えられ、情報を構造化することのさらなる意義が示唆された。今後更なる検討を行う予定である。

参考文献

- 1) 平成29年度情報通信白書，総務省
- 2) Cancer research UK. About Cancer Research UK's patient and health information - Writer guidelines- - Researching and writing-. <http://www.cancerresearchuk.org/about-cancer/about-our-information/writers-guidelines/researching-writing#collapse-207365>
- 3) 後藤あや. (2015). ヘルスリテラシー-健康に関する情報を使う力・伝える力-. Isotope News No. 732.

F. 健康危険情報

なし

G. 研究発表

I 著書 なし

II 総説 なし

III 原著 なし

IV 症例報告 なし

V 学会発表 なし

H. 知的財産権の出願・登録状況

1. 特許取得
なし
2. 実用新案登録
なし
3. その他
なし

附表1 構成要素一覧

番号	構成	見出し	項目	①共通 ②がん種特有 ③削除または 統合× ④追加☆	ピース 数	⑤冊子共通◎ ⑥冊子がん種特 有△ ⑦冊子では不使 用×	ピース 数	
1		がんの診療 の流れ		×		◎		
2		目次		×		◎		
3		がんと言われ たあなたの心 に起こる こと		×		◎		
4	基礎 知識	●●について	臓器の構造	◎	1	◎	1	
5			臓器の機能	◎	2/3	◎	2/3	
6			臓器に付随するもの(内分泌な ど)の正常な機能	△	2	×		
7		●●がんと	がんの原発部位	◎	1	◎	1	
8			がんの転移部位	△	1	×		
9		症状	自覚症状概要	◎	2/3	◎	2/3	
10			自覚症状	△	5	△	5	
11			早期受診のすすめ	△	1	×		
12		病型分類/組 織型分	病型分類概要	△	2/3	△	2/3	
13			病型分類	△	3 1/3	△	3 1/3	
14		関連する疾 患	関連する疾患	△	1 1/3	×		
15		統計	罹患率(年代別、性別、部位 別)	◎	1	◎	1	
16		発生要因	発生要因	◎	1	◎	1	
17			発生要因	△	1	△	1	
18		予防と検診	予防	◎	1	×		
19			検診	◎	1	×		
20			検診の内容	△	1	×		
21		「●●がん」 参考文献	参考文献	◎	1	×		
22		検査	●●がんの 検査	検査概要	◎	2/3	◎	2/3
23			検査の種類	検査種類	△	7	△	7
24	治療	病期と治療の 選択	病期と治療の選択概要	◎	2/3	◎	2/3	
25			各種がんの病期分類	◎	2/3	◎	2/3	
26			病期から推測する治療方針/早 期がんと進行がん	△	1	△	1	
27			深達する粘膜詳細	△	1	△	1	
28			深達度	△	1	△	1	
29			サブタイプ分類	△	1	△	1	
30			予後因子とは	△	2/3	×		
31			よくない経過の病期	△	1	×		
32			予後因子を考慮した治療	△	1	△	1	
33			治療方法概要/治療の時期	◎	2/3	◎	2/3	
34			アルゴリズムの概要	◎	2/3	◎	2/3	
35			アルゴリズムの説明	△	2	×		
36			アルゴリズムの説明	△	2	×		
37			妊娠・出産について	△	1	×		

番号	構成	見出し	項目	①共通 ②がん種特 有 ③削除また は統合× ④追加☆	ピース 数	⑤冊子共通◎ ⑥冊子がん種 特有△ ⑦冊子では不 使用×	ピース 数
38	治療	内視鏡治療	内視鏡治療の概要(目的/方法)	△	1	△	1
39			内視鏡治療の適応の詳細	△	2	△	2
40		手術(外科治療)	手術の概要	◎	2/3	◎	2/3
41			手術方法(部位別)	◎	2	◎	2
42			術式、その他	△	6	△	6
43			術直後の様子	△	1	×	
44			術後合併症概要	◎	2/3	◎	2/3
45			術後合併症	◎	3	◎	3
46			放射線治療	放射線治療概要/方法	◎	1	◎
47		放射線治療の詳細		◎	2	◎	2
48		放射線治療の詳細2		◎	2	◎	2
49		薬物療法	薬物療法の概要	◎	2/3	◎	2/3
50			それぞれの療法の概要1	◎	1	◎	1
51			薬剤名、副作用、その他	◎	1	×	
52			薬剤の種類、副作用、その他	◎	4	◎	4
53			それぞれの療法の概要2	△	1	△	1
54			薬剤名、副作用、その他	△	1	×	
55			薬剤の種類、副作用、その他	△	4	△	4
56		生存率	生存率について	◎	1	×	
57			公表している生存率/5年生存率	◎	1	×	
58		リハビリテーション	リハビリテーション概要	☆	1	×	
59			リハビリテーション詳細	☆	3	×	
60		合併症に対する治療/支持療法	治療概要	△	2/3	×	
61			合併症に対する治療/支持療法	△	5	×	
62		緩和ケア	緩和ケアとは/緩和ケアを始める時期/目的と方法	◎	1	×	
63		臨床試験	臨床試験とは	◎	1	×	
64			●●がんの臨床試験	◎	1 1/3	×	
65		転移・再発	転移・再発とは	◎	2/3	◎	2/3
66			転移好発部位/転移後の治療	◎	1 1/3	◎	1 1/3
67			各種癌の再発	◎	1 2/3	◎	1 2/3
68	再発や転移を知る/		◎	1	×		
69	療養	日常生活を送る上で	日常生活の心配事/注意点	◎	1	×	
70			主な後遺症への対策概要	△	2/3	×	
71			主な後遺症への対策	△	5	×	
72		性生活・妊娠について	◎	1	×		
73	経過観察	術後の定期検査	◎	1	◎	1	
74		術後の定期検査内容	◎	1	◎	1	
75		診断や治療の方針に納得できましたか?	×		◎		
76		メモ	×		◎		
			目安の文字数		10800		10400