

2020-A-19 がん情報提供の自動化に向けた研究

八巻知香子

がん対策研究所 がん情報提供部 患者市民連携推進室

研究の分野名

社会医学・国際戦略分野

研究の概要

第3期がん対策推進基本計画では、医療技術や情報端末が進歩し、がん患者の療養生活が多様化する中で、患者やその家族の治療上の疑問や、精神的・心理社会的な悩みも多様化していることから、拠点病院等や小児がん拠点病院のがん相談支援センターが中心となって相談支援の一層の充実をすべきことがあげられた。また、氾濫しているインターネット等での情報には、真偽の確かでない情報が含まれていることがあり、正しい情報を得ることが困難な場合があることも指摘されたように、患者や家族の疑問や不安も増大している。がん情報サービスでは、各種診療ガイドライン等のエビデンスレベルの高い情報を基に情報を作成し、提供しているが、今後も増え続けると考えられるがんや医療に関する情報に対して、患者や家族をはじめとする国民が、がんに関する正しい情報をより迅速に適切に活用できるようにするためのしくみが求められている。

そこで、本研究では、信頼を確保しつつ効率的にがんの情報をより多くの人に届けるために、新しい情報技術を用いた情報提供の自動化について検討する。

将来的には、人工知能による音声対話による相談が理想だが、その情報基盤となるがん患者・家族の情報ニーズは集積されていない。さらに、がんに関する相談は、多種多様であり、医療以外の一般製品のコールセンターでの相談対応のように、容易に一定の質問と回答に集約することは難しく、大量のデータを解析することが必要となるが、その手法は検討されていない。そこで、本研究では、昨年までに行った(1)大量のデータより患者・家族の情報ニーズを収集し、解析する手法の検討の内容をもとに、(2)チャットによるがん相談システムの導入手法の検討を目的とする。

研究経費

8,383千円(繰越分含む)

研究班の組織

研究者名	所属研究機関名・職名	分担研究課題名
(研究代表者) 八巻知香子	国立研究開発法人国立がん研究センターがん対策研究所がん情報提供部患者市民連携推進室 室長	チャットでの相談システムの検討
早川 雅代	国立研究開発法人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 外来研究員	音声認識システム・テキストマイニングの検討、チャットでの相談システムの検討
高山 智子	国立研究開発法人国立がん研究センターがん対策研究所がん情報提供部 部長	音声認識システム・テキストマイニングの検討、チャットでの相談システムの検討

内村 祐之	東京医科歯科大学 東京医科歯科大学病院 医療情報部 特任講師	音声認識システム・テキストマイニングの検討、チャットでの相談システムの検討
安齋 達彦	東京医科歯科大学 M&Dデータ科学センター 講師	音声認識システム・テキストマイニングの検討、チャットでの相談システムの検討
花出 正美	がん研究会有明病院 がん相談支援センター 看護師長	チャットでの相談システムの検討
高橋 朋子	国立研究開発法人国立がん研究センターがん対策研究所がん情報提供部がん相談支援推進室 看護師	チャットでの相談システムの検討
渡邊 乙女	国立研究開発法人国立がん研究センターがん対策研究所がん情報提供部情報高度化企画室 研究員	音声認識システム・テキストマイニングの検討
櫻井 雅代	国立研究開発法人国立がん研究センターがん対策研究所がん情報提供部がん相談支援推進室 看護師	チャットでの相談システムの検討
志賀 久美子	国立研究開発法人国立がん研究センターがん対策研究所がん情報提供部がん相談支援推進室 看護師	チャットでの相談システムの検討
佐藤 稔子	国立研究開発法人国立がん研究センターがん対策研究所がん情報提供部がん相談支援推進室 薬剤師	チャットでの相談システムの検討
佐野 由美子	国立研究開発法人国立がん研究センターがん対策研究所がん情報提供部がん情報企画室 研究員	音声認識システム・テキストマイニングの検討
三村 麻子	国立研究開発法人国立がん研究センターがん対策研究所がん情報提供部患者市民連携推進室 研究員	音声認識システム・テキストマイニングの検討
石川 文子	国立研究開発法人国立がん研究センターがん対策研究所がん情報提供部がん情報企画室 室長	音声認識システム・テキストマイニングの検討
堀抜 文香	国立研究開発法人国立がん研究センターがん対策研究所がん情報提供部たばこ政策情報室 研究員	音声認識システム・テキストマイニングの検討、チャットでの相談システムの検討

宮本 紗代	国立研究開発法人国立がん研究センターがん対策研究所がん情報提供部がん相談支援推進室 看護師	チャットでの相談システムの検討
齋藤 睦美	国立研究開発法人国立がん研究センターがん対策研究所がん情報提供部患者市民連携推進室 非常勤研究員	チャットでの相談システムの検討
羽山 慎亮	国立研究開発法人国立がん研究センターがん対策研究所がん情報提供部患者市民連携推進室 非常勤研究員	テキストデータのテキストマイニング検討

研究の目的と到達目標及び実績要点

全期間

（目的と到達目標）

（目的）

本研究では、信頼を確保しつつ効率的にがんの情報をより多くの人に届けるために、新しい情報技術を用いた情報提供の自動化について検討することを目的とする。具体的には、（１）大量のデータより患者・家族の情報ニーズを収集し、解析する手法の検討及び（２）チャットによるがん相談システムの導入手法の検討を行う。将来的に、これらの仕組みで集積するデータが大規模となり、患者や家族等への人工知能による音声対話による相談に向けた基盤となることを目指す。

（到達目標）

１）大量のデータより患者・家族の情報ニーズを収集し、解析する手法の検討

先行研究班「わが国におけるがん情報データベースのあり方と評価に関する研究（29-A-18）」において取得した7がん種（乳がん、肺がん、大腸がん、血液がん、すい臓がん、胃がん、前立腺がん）の電話相談の音声のテキストデータおよび、２）でのチャットによるがん相談により蓄積される同がん種のテキストデータから、機械学習（AI技術の1種；Wordvec2を使用）による抽出型要約技術を用いて、大量のデータより患者・家族の情報ニーズを抽出し、病院探しをしているのか、治療の選択で悩んでいるのか、不安の相談なのかなどに応じた的確な情報の提示ができるための糸口を探索する。
また、解析データを用いて、媒体別の相談利用者を考慮して的確な情報を提示するための素材としてのFrequent asked question（FAQ）を大量に作成するための手法の検討を行う。

２）チャットによるがん相談システムの導入手法の検討

チャットによるがん相談システムとは、相談員（人）がPC上で入力して相談対応するシステムを想定している。システム要件を洗い出し、市販のチャットシステムの導入とカスタマイズを軸に仕様を検討する。がん情報サービスサポートセンター及びがん相談支援センター数施設にて導入を試みる。
チャットでの相談対応では、電話相談での対応とは異なるスキルと留意点があることが想定されることから、導入前にマニュアルを作成し、実際の導入後にその評価と改良を行う。
また、電話相談（音声相談記録）とチャット相談（テキスト相談記録）の相談の特徴（利用者属性、がん種、相談対応時間、相談内容等）を比較することにより、チャット相談のニーズを把握し、導入の可否の検討につなげる。

【1年次計画計画提出後の変更点と変更の理由について】

本研究の1年目であるR2年は、新型コロナウイルス感染症の影響により、急速にオンラインでの生活環境の導入・整備が国内で進むという事態となった。研究実施主体である国立がん研究センターがん対策情報センターがん情報サポートセンターにおいても、出勤が困難な状況が続き、電話相談対応への影響があった。

また今後も同様な状況が発生することも懸念された。以上より、1) 大量のデータより患者・家族の情報ニーズを収集し、解析する手法の検討については、計画立案時当初（R2 年 1 月）の、電話相談をもとにした音声相談記録を主とする情報ニーズの蓄積を図る手法の検討から、本研究計画の 2) において検討を進めるチャット相談による大量のデータ蓄積を行う方針へ計画変更を行った。この方針変更に伴い、前年度継続（開発費 29-A-18）の乳がん他全 7 がん種の相談記録とチャット相談記録を解析対象として、大量のデータより機械的に情報ニーズを抽出する手法を探るとともに、電話相談（音声相談記録）とチャット相談（テキスト相談記録）の相談の特徴（利用者属性、がん種、相談対応時間、相談内容等）を明らかにすることとした。また、解析データを用いて、媒体別の相談利用者を考慮して的確な情報を提示するための素材となりうる **Frequent asked question (FAQ)** を大量に作成するための手法の検討を行うこととした。以上により、下線部にあげるように、到達目標の変更が生じた。

当該年次

（到達目標）

（3 年次到達目標） ※800 字以内

1) 大量のデータより患者・家族の情報ニーズを収集し、解析する手法の検討

前年度の検討をもとに、媒体別の相談利用者を考慮して的確な情報を提示するための **FAQ** を選別し、チャットでの実運用により評価する。また、評価結果より、チャットボットシステムの構築への課題を抽出する。

2) チャットによるがん相談システムの導入手法の検討

利用者記録を用いた解析を行い、患者・家族の情報ニーズを解析する。また、解析結果より、チャットボットシステムの構築への課題を抽出する。

【1 年次計画計画提出後の 3 年次の到達目標の変更の理由について】

変更の理由は、1 年次の到達目標の変更の理由と同様。また 1 年次の変更時点に 3 年次の到達目標の変更を行った。

（年次評価時点の実績要点）

1) 大量のデータより患者・家族の情報ニーズを収集し、解析する手法の検討

＜電話相談データの活用＞

A：音声データを文字データに変換する、B：文字データの質問を抽出する、C：文字データの質問に適切な回答を結びつけるという 3 段階を検討した。

「A：音声データを文字データへの変換」については、医療分野での日本語音声認識エンジンのシェア 1 位である、「AmiVoice」を保有するアドバンスト・メディア社との共同研究契約を結び、「共同研究用音声認識辞書」の構築を実施した。実際の電話相談（匿名化加工済み）とその逐語録を学習データとして用いることにより、既存の音声認識辞書では 78.08%の認識率であったが、本共同研究用音声認識辞書では 86.34%まで向上した。音声認識による文字化は十分な効果が見込めると判断できた。

「B：文字データの質問を抽出する」「C：文字データの質問に適切な回答を結びつける」手法については **Word2Vec** を用いた自動化を試みたが現時点での自動化は困難と判断し、断念した。

以上より、電話相談データについては、音声を文字認識する段階までは実用の可能性が高いが、自動的に **FAQ** を作成することは現時点では困難と判断された。

＜チャット相談データの活用＞

研究計画変更に伴い導入することとなったチャット相談のデータを用いて、①チャット相談の会話パターンの把握、②チャットボットシステムの試作と、回答の自動化の可能性についての検討を実施した。対人チャット相談データを学習用データとして用い、回答を自動生成する機能「**AI 型チャットボット**」を試作したうえで、実用可能性について、有人相談を担当する相談員 4 名を交えて検討した。ある程度有意義な回答を返すこともでき、多量の学習データを提供すれば、実用可能なものになる可能性が示唆された。

が、不適切な回答を返すことのリスクは排除できない。そのため、「AI 型チャットボット」は、相談の最終的な回答を示す機能としてではなく、応答の中で質問を絞り込むために利用し、最終的には必要な情報のリンク先の提示（有人相談に繋げる選択肢を含む）に繋げる「シナリオ型チャットボット」で回答を返すことを目指すことが、近い将来に一部の自動化を達成するうえで有用な道筋となることの見込みがたつた。

2) チャットによるがん相談システムの導入手法の検討

文字による相談と回答のデータとして利用可能な状況となった。2020 年度（12～1 月）には 27 件／月、2021 年度上期には 24 件／月であった相談件数は、下期 30.1 件、2022 年度上期 35.5 件／月、下期（10～11 月）41.5 件／月に増加してきている。研究に利用できるデータとして、開始から 2022 年 11 月末までに計 741 件のチャット相談データが蓄積された。

チャット相談は、電話相談と共通した特徴をもちつつも、異なったニーズを捕捉できていること、異なる利点を持ちうるということが明らかになり、そのニーズをつかみ、適切に対応することで、相談方法の選択肢の広がりや、相談内容の質そのものの向上にも寄与しうるものと考えられた。

1) に記載の通り、チャットボットの学習データとして有用であり、将来展開としては、電話相談、チャットボットによる情報提示と相補的な役割を果たすことが期待されると考えられた。

研究成果と考察

当該次時評価時点

1) チャットによるがん相談システムの導入手法の検討

研究計画の変更により開始したチャット相談が軌道にのり、文字による相談と回答のデータとして利用可能な状況となった。2020 年度（12～1 月）には 27 件／月、2021 年度上期には 24 件／月であった相談件数は、下期 30.1 件、2022 年度上期 35.5 件／月、下期（10～11 月）41.5 件／月に増加してきている。研究に利用できるデータとして、開始から 2022 年 11 月末までに計 741 件のチャット相談データが蓄積された。

チャット相談の利用数は徐々に増え、実際に相談者を支援することによる効用だけでなく、2) で後述するとおり、患者や家族等のニーズに沿った情報提示に結び付けるためのチャットボットの学習データとして有用である可能性が高い。今後の継続的な運営とデータの特性を検討するにあたり、チャット相談導入のコスト（導入や実施に必要な準備）、チャット相談の特徴とそれに応じた相談スキルについて、実際に相談対応を行っている相談員へのフォーカスグループインタビューにより検討した。

2022 年 3 月に、相談に対応する 4 名の相談員にディスカッションによるディスカッション内容を「①チャット相談導入や実施に必要な準備」「②チャット相談の特徴」「③相談員がチャット相談で活用するスキル」「④チャット相談の展開に向けて必要なこと」の観点で整理し、該当する語りを抽出し、抽出した語りに、内容を端的に表す短い文章をつける（コーディング）、①～④別に、類似した内容のコードをまとめ、カテゴリを作成する（カテゴリ化）、作成したカテゴリを用いて、チャット相談の特徴や求められる姿をまとめた。

「①チャット相談の準備」としては、チャットシステムについての情報収集と選定、相談マニュアルの整備が必要であったと指摘された。

「②チャット相談の特徴」としては、「（電話相談に比べて）相談者のペースに沿いやすい」「相談者の年齢、仕様機器、タイピング速度に幅がある」「相談者の周囲の状況にかかわらず相談しやすい」「書くこと」「文字に残ること」が相談者・相談員にもたらすものがある」「URL のやり取りで情報にリンクできる URL のやり取りで情報にリンクできる」といった特徴が挙げられた。

「③相談員が活用する相談スキル」としては、「電話相談のスキルを基礎にする」「限られた言葉や文字数で共感を示し、相談者のニーズにたどり着く」「相談者の文字入力タイミングや文体の雰囲気を捉える」「相談者の入力時間に立ち入らない」「相談全体の流れを描き、意図的に相談者への返信を組み立てる」「相談者に返信する文章の雰囲気を和らげ、人間味を持たせる」「情報提供と返信のための下準備をしておく」が挙げられた。

「④今後のチャット相談の展開」としては、「チャット相談特有のニーズに応える」「ハードを整備する必要がある」「情報提供はチャットボットが、意思決定支援はチャット相談が担う体制を整備する」ことが必

要であるとの意見が述べられた。

以上より、相談員には、チャット相談が電話相談と共通した特徴をもち、そのスキルを応用しながら対応していることを自覚していること、ただ、チャット相談の方が適した相談もあることを認識し、その利点を生かしつつ相談者に合わせた対応を行っていることが明らかとなった。

2) 大量のデータより患者・家族の情報ニーズを収集し、解析する手法の検討

3年時の目標は、前年度の検討をもとに、①媒体別の相談利用者を考慮して的確な情報を提示するためのFAQを選別し、チャットでの実運用により評価すること、②評価結果より、チャットボットシステムの構築への課題を抽出することであった。

(1) 電話相談データの利用可能性の検討

1年次、2年次には、A：音声データを文字データに変換する、B：文字データの質問を抽出する、C：文字データの質問に適切な回答を結びつけるという3段階を検討した。

「A：音声データから文字データへの変換」は、2年次時までの検討結果を踏まえ辞書機能の向上が必須であると判断した。音声認識は、使われる環境に適した辞書を登用することで実用可能なレベルの認識が可能となる。医療分野での日本語音声認識エンジンのシェア1位である、「AmiVoice」を保有するアドバンスト・メディア社との共同研究契約を結び、「共同研究用音声認識辞書」の構築を実施した。

匿名化した電話相談の音声7通話を学習データとし、音響モデルとしては、コールセンター系自然発話用の音響モデルを採用した。

また、匿名化した電話相談の逐語録217通話をを用いて、がん情報サービスサポートセンターコーパスを作成した。そのうえで、言語モデルとしては、「汎用（一般用語）4：国立がん研究センター受領コーパス4：カルテ記載用コーパス3」の割合で学習を実施した。

3件の通話（逐語録および音声）を評価データとして検証した結果、アドバンスト・メディア社が保有する既存の音声認識辞書では78.08%の認識率であったが、本共同研究用音声認識辞書では86.34%まで向上した。

この結果を鑑み、チャットボット開発のためのテキストデータの書き起こしについて、音声認識による文字化は十分な効果が見込めると判断した。

「B：文字データの質問を抽出する」ための方策として、1・2年次に、電話相談の逐語録データを素材として、Word2Vec技術を用いた機械学習により多量のFAQを作成する方法の可能性を検討した。しかし、既存の技術では、機械抽出した重要文を要約作成に適用しても【見逃し】の割合が高いこと、抽出された重要文の約半数弱が【ノイズ】であることから、機械抽出した重要文を採用することは効果的とはいえない状況であった。

「C：文字データの質問に適切な回答を結びつける」手法については、前述のWord2Vecを用いて、手動で作成したFAQを学習させる手法について検討を試みたが現時点での自動化は困難と判断した。

2022年の段階では、電話相談の録音データから、B：文字データの質問を抽出する、C：文字データの質問に適切な回答を結びつける、工程をこなすことを実用化するのは困難であり、文字データで相談が展開されるチャット相談データに重きを置くことが現実的であると判断された。

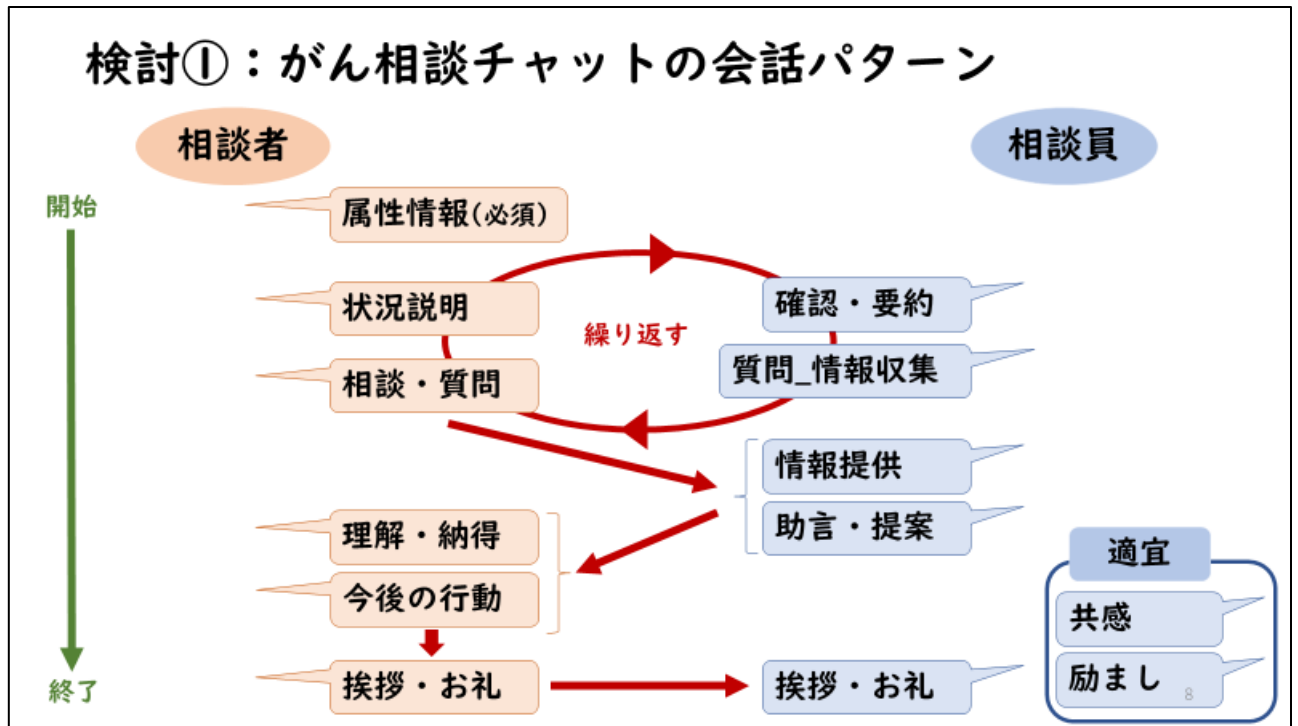
(2) チャット相談データを用いたチャットボットの可能性の検討

1で導入したチャット相談のデータを用いて、①チャット相談の会話パターンの把握、②チャットボットシステムの試作と、回答の自動化の可能性についての検討を実施した。

① チャット相談の会話パターンの把握

チャット相談の発言は、相談者による「状況説明」「相談・質問」「理解・納得」「今後の行動」、相談員による「確認・要約」「情報収集のための質問」「情報提供」「助言・提案」「共感」「励まし」、双方による「挨拶・お礼」「その他」に大別された。

これらの発言は、相談者による属性情報の入力のもと、相談者からの「状況説明」「相談・質問」の提示、相談員による「確認・要約」と、「情報収集のための質問」が行われ、さらに相談者からの「状況説明」「相談・質問」が引き出されるという流れを繰り返したのち、「情報提供」「助言・提案」がなされ、相談者から「理解・納得」「今後の行動」が述べられ互いに「挨拶・お礼」が提示されるという流れとなっていた。



図：チャット相談のパターン

②チャットボットシステムの試作と、回答の自動化の可能性についての検討

学習用データとして用いた 233 相談、5132 発話のうち、相談者から発せられた「相談・質問」は 629 発話あり、そのうちチャットボットが回答できる可能性があるのは、そのうち 113 発話（18.0%）であった。個別性の高い相談、医学的判断が求められる相談、漠然とした問いかけについては対人チャットが望ましいと考えられた。

過去のデータをもとに回答を自動生成する機能「AI 型チャットボット」を試作したうえで、実用可能性について、有人相談を担当する相談員 4 名を交えて検討した。様々な質問を試入力してみたところ、ある程度有意義な回答を返すこともでき、多量の学習データを提供すれば、実用可能なものになる可能性が示唆された。しかしながら、がんについて不安や悩みを入力した際に、不適切な回答を返すことのリスクは排除できない。そのため、「AI 型チャットボット」は、相談の最終的な回答を示す機能としてではなく、応答の中で質問を絞り込むために利用し、最終的には必要な情報のリンク先の提示（有人相談に繋げる選択肢を含む）に繋げる「シナリオ型チャットボット」で回答を返すことを目指すことが、近い将来に一部の自動化を達成するうえで有用な道筋となることの見込みがたつた。

倫理面への配慮

(1) 遵守すべき研究に関する指針等

- ☐ 再生医療等の安全性の確保等に関する法律
- ☐ 臨床研究法
- ☐ 医薬品の臨床試験の実施に関する基準（GCP）
- ☒ 人を対象とする医学系研究に関する倫理指針
- ☐ ヒトゲノム・遺伝子解析研究に関する倫理指針
- ☐ 遺伝子治療臨床研究に関する指針
- ☐ 動物実験等の実施に関する基本指針

☒ その他の指針等（指針等の名称：人を対象とする生命倫理・医学系研究に関する倫理指針）

(2) 本研究開発期間中に予定される臨床研究の有無

☐ 有

☒ 無

本研究に関連する、本研究期間中の主な発表論文等

当該年次

(学術誌論文)

- ・国立がん研究センター研究開発費による成果であることが記載されているもの

Hayakawa M, Watanabe O, Shiga K, Fujishita M, Yamaki C, Ogo Y, Takahashi T, Ikeguchi Y, Takayama T. Exploring types of conversational agents for resolving cancer patients' questions and concerns: Analysis of 100 telephone consultations on breast cancer. Patient Education and Counseling. 2022. Oct 10;S0738-3991(22)00447-5. doi: 10.1016/j.pec.2022.10.004. Online ahead of print. (査読有)

- ・国立がん研究センター研究開発費による成果であることが記載はないが、関連するもの

記載なし

(学会発表)

八巻知香子, 清水奈緒美, 小郷祐子, 櫻井雅代, 齋藤弓子, 堀抜文香, 高山智子. がん専門相談員を対象とする「情報支援研修」の効果測定に関する研究. 第 60 回日本癌治療学会学術集会.2022.10.20-22. 神戸.

堀抜文香, 八巻知香子, 井上洋士, 池口佳子, 高山智子. がん相談支援センターを初めて利用する相談者の利用前後の心理状態や生活の質 (QOL) に関する検討. 第 7 回日本がんサポーターティブケア学会学術集会. (ハイブリッド開催) (2022. 6).

堀抜文香, 齋藤弓子, 石川文子, 佐野由美子, 高山智子, 若尾文彦. がんの情報入手と e ヘルスリテラシーとの関連の検討: がん情報サービス利用者調査から. 第 60 回日本癌治療学会学術集会. (神戸) (2022. 10)

櫻井雅代, 佐藤稔子, 志賀久美子, 澤井映美, 宮本紗代, 八巻知香子, 高山智子. COVID-19 の流行前後におけるがん電話相談の変化とがん患者や家族の経験. 第 10 回 日本がん相談研究会年次大会.2022.3.12. 東京.

(書籍)

記載なし

(知的財産権)

記載なし

(政策提言 (寄与した指針等))

記載なし

(その他)

記載なし