



مرکز آموزش عالی
علوم پزشکی
وارستگان

به نام خدا

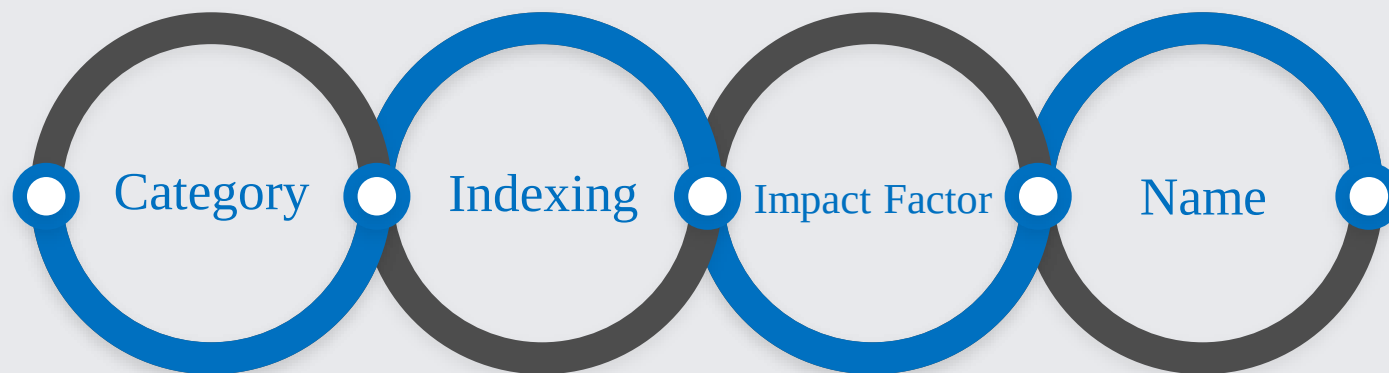
An assessment of ChatGPT's responses to frequently asked questions about cervical and breast cancer

ارزیابی پاسخ های ChatGPT به سوالات رایج درباره سرطان دهانه رحم و سرطان سینه

JOURNAL

ISI, Scopus, PubMed
Embase, DOAJ

BMC Women's Health



Q2

2.400

اختصارات

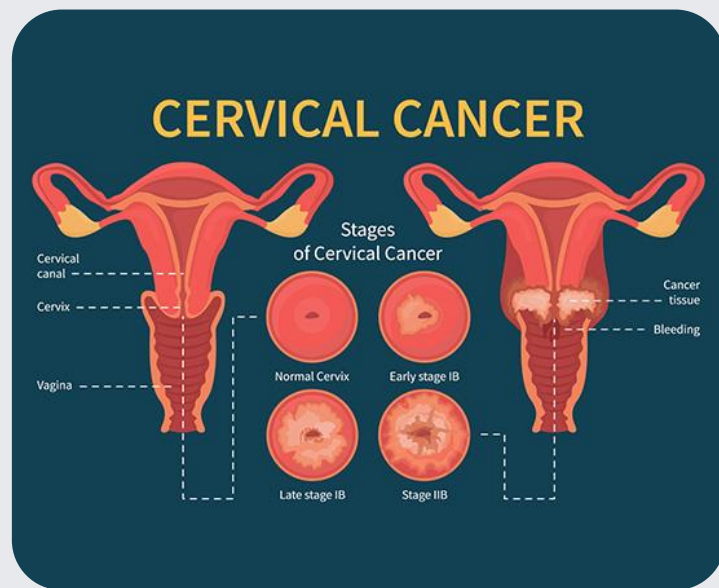
ABBREVIATION	اصطلاح کامل	توضیح
HPV	Human Papilloma Virus	ویروس پاپیلومای انسانی
HDI	Human Development Index	شاخص توسعه انسانی
CC	Cervical Cancer	سرطان دهانه رحم
BC	Breast Cancer	سرطان پستان
FKGL	Flesch-Kincaid Grade Level	شاخص استاندارد برای ارزیابی خوانایی و سادگی متون و تعیین این که یک متن برای چه گروه سنی یا سطح تحصیلی مناسب است.
FRES	Flesch Reading Ease Score	شاخص استاندارد برای ارزیابی میزان دشواری یک متن که آن را بر اساس مقیاس عددی اندازه گیری می کند.
NLP	Natural Language Processing	پردازش زبان طبیعی که به زبان ساده یعنی یاد دادن زبان انسانی به کامپیوتر برای درک زبان انسان
GPT	Generative Pre-trained Transformer	یک مدل هوش مصنوعی که توسط شرکت OpenAI توسعه داده شده است و هدف آن تولید و پردازش زبان طبیعی است.



سرطان دهانه رحم

- چهارمین سرطان شایع در زنان در سراسر جهان
حدود ۶۶۰ هزار مورد جدید در سال ۲۰۲۲
حدود ۳۵۰ هزار مرگ و میر در سال ۲۰۲۲
بیشترین میزان مرگ و میر در کشورهای با درآمد کم یا متوسط (۹۴٪)

کشورهای جنوب آفریقا
آمریکا مرکزی
آسیا جنوب شرقی



• عامل آن HPV

- عوامل خطر: نوع ویروس، وضعیت ایمنی، تعداد تولد، سن پایین در بارداری اول، داروهای ضدبارداری هورمونی و سیگار

• قابل درمان در مراحل اولیه

- پیشگیری: واکسیناسیون، وسایل پیشگیری، غربالگری و درمان ضایعات پیش سرطانی

- علائم آن شامل خونریزی غیرطبیعی واژینال، درد لگن و ترشحات غیر معمول

مقدمه و
بیان مسئله

فهرست

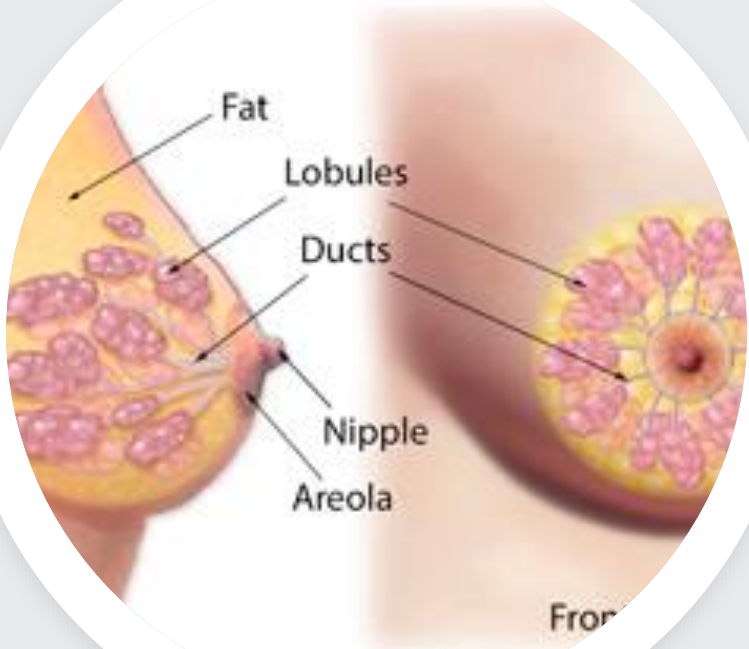
سرطان پستان

Ductal carcinoma

Lobular carcinoma

Inflammatory breast cancer

Paget's disease of the breast



فهرست

مقدمه و
بیان مسئله

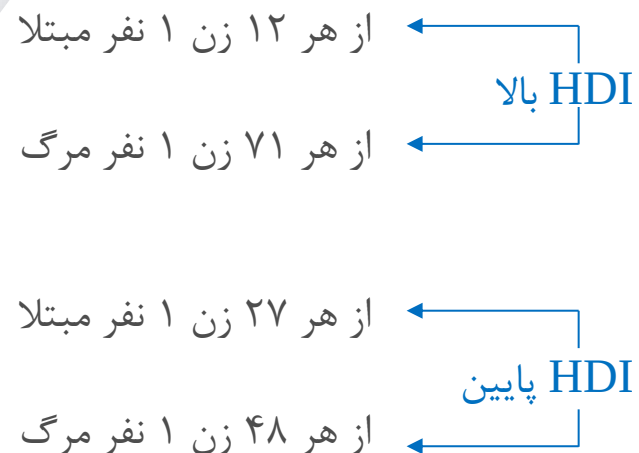
۱

سرطان پستان

- شایع ترین سرطان در زنان ۱۵۷ کشور از ۱۸۵ کشور
۲/۳ میلیون زن مبتلا در سال ۲۰۲۲
حدود ۶۷۰ هزار نفر مرگ و میر در سال ۲۰۲۲
۰/۵ تا ۱ درصد از آن ها در مردان

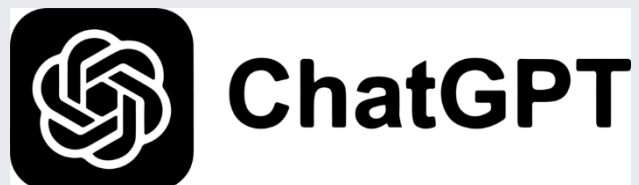
عوامل خطر: جنسیت، افزایش سن، چاقی، سوء مصرف الکل، سابقه خانوادگی سرطان سینه، سابقه قرار گرفتن در معرض اشعه، سابقه باروری (مانند سن شروع قاعدگی و سن در اولین بارداری)، مصرف دخانیات و هورمون درمانی پس از یائسگی

درمان: جراحی (لامپکتومی یا ماستکتومی)، پرتودرمانی، دارو های هورمونی و شیمی درمانی



ChatGPT

- یک مدل زبان هوش مصنوعی که توسط OpenAI توسعه یافته است.
- بر اساس معماری GPT که یک NLP است.
- توانایی تولید و درک زبان انسانی را به صورت متنی دارد.



اهداف

مقایسه خروجی های ChatGPT با پاسخ های پزشکان به سوالات متداول درباره سرطان دهانه رحم و سرطان پستان



تجزیه و تحلیل شاخص های ارزیابی ثانویه:

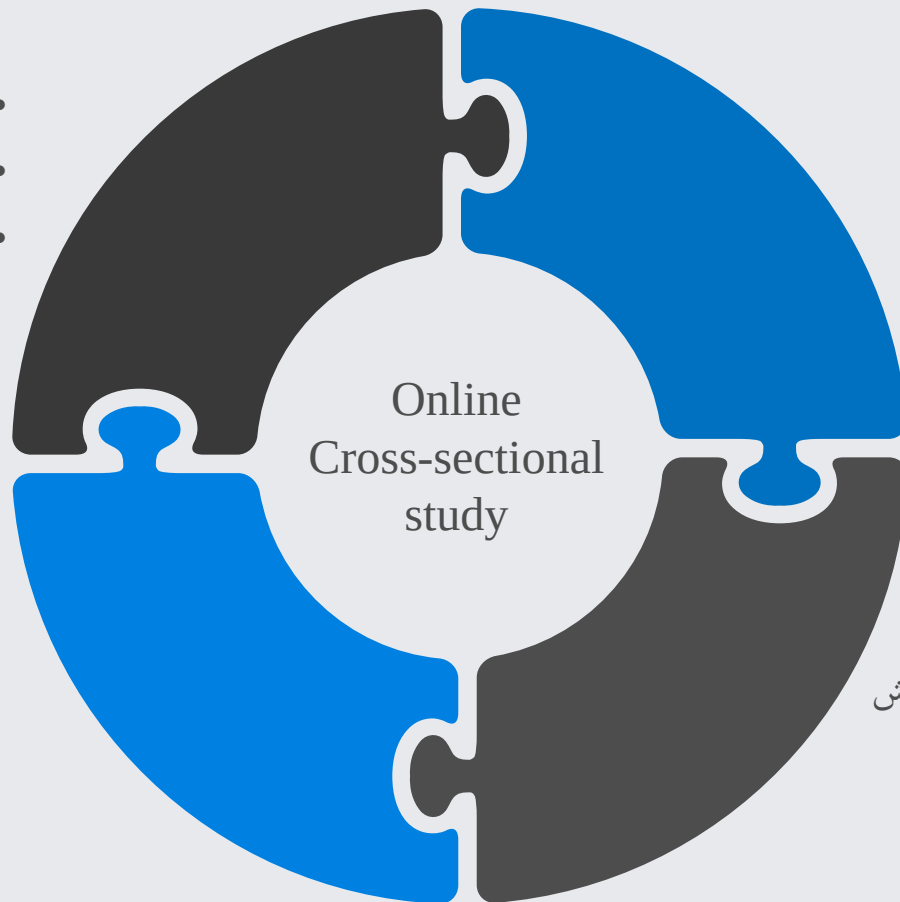
- ✓ دقت
- ✓ خوانایی
- ✓ سازگاری
- ✓ قابلیت اطمینان
- ✓ کارایی
- ✓ ارتباط
- ✓ محتوای پاسخ ها



ارزیابی پاسخ های ChatGPT به سوالات متداول سرطان دهانه رحم و سرطان پستان

طراحی مطالعه و شرکت کنندگان

- پنج متخصص ماموگرافی
- پنج متخصص زنان
- ۲۰ سوال رایج علمی



مارس تا می ۲۰۲۳

استاندارد مرجع :

- اپیدمیولوژیست های مجرب تومور با بیش از ۲۰ سال سابقه در این دو سرطان
- دو مامولوژیست
- دو متخصص زنان

نمونه های ثبت شده در
شیلی و آلمان در ۲۶ و
۲۹ آوریل ۲۰۲۳



ارزیابی پاسخ های ChatGPT به سوالات متداول سرطان دهانه رحم و سرطان پستان

مراحل

مرحله	عنوان	توضیحات
۱	فرمول بندی سوالات و پاسخ های «استاندارد مرجع»	برای هر سرطان ۱۰ سوال طراحی شد و متخصصان به سوالات حوزه خود پاسخ دادند. همچنین از کارشناسان خواسته شد که سوالات را بررسی و پاسخ های اجماع کارشناسان را تدوین کنند. مجموعه نهایی این پاسخ ها با چندین دور مشاوره و ارزیابی تعیین شد.
۲	جمع آوری پاسخ ها	دو محقق در زمان ها و مکان های مختلف به ChatGPT 3.5 دسترسی پیدا کردند و ۲۰ سوال از پیش تعیین شده را با استفاده از یک قالب استاندارد مطرح کرده و هر یک را به طور مکرر سه بار پرسیدند. همچنین منابع پاسخ ها از ChatGPT خواسته شد.
۳	ارزیابی پاسخ ها	پاسخ ها به طور تصادفی مرتب شدند و کارشناسان برای هر سوال، ۶ پاسخ را ارزیابی کردند که در این ارزیابی دقت و سازگاری پاسخ ها مورد توجه بود.
۴	ارزیابی جامع	عملکرد ChatGPT به طور بصری با یک نمودار راداری ارزیابی شد. این نمودار دقت، خوانایی، سازگاری، قابلیت اعتماد، کارایی و ارتباط را ترکیب کرد تا قابلیت استفاده از ChatGPT در پیشگیری و درمان جامع CC و BC را منعکس کند.



ارزیابی پاسخ های ChatGPT به سوالات متداول سرطان دهانه رحم و سرطان پستان

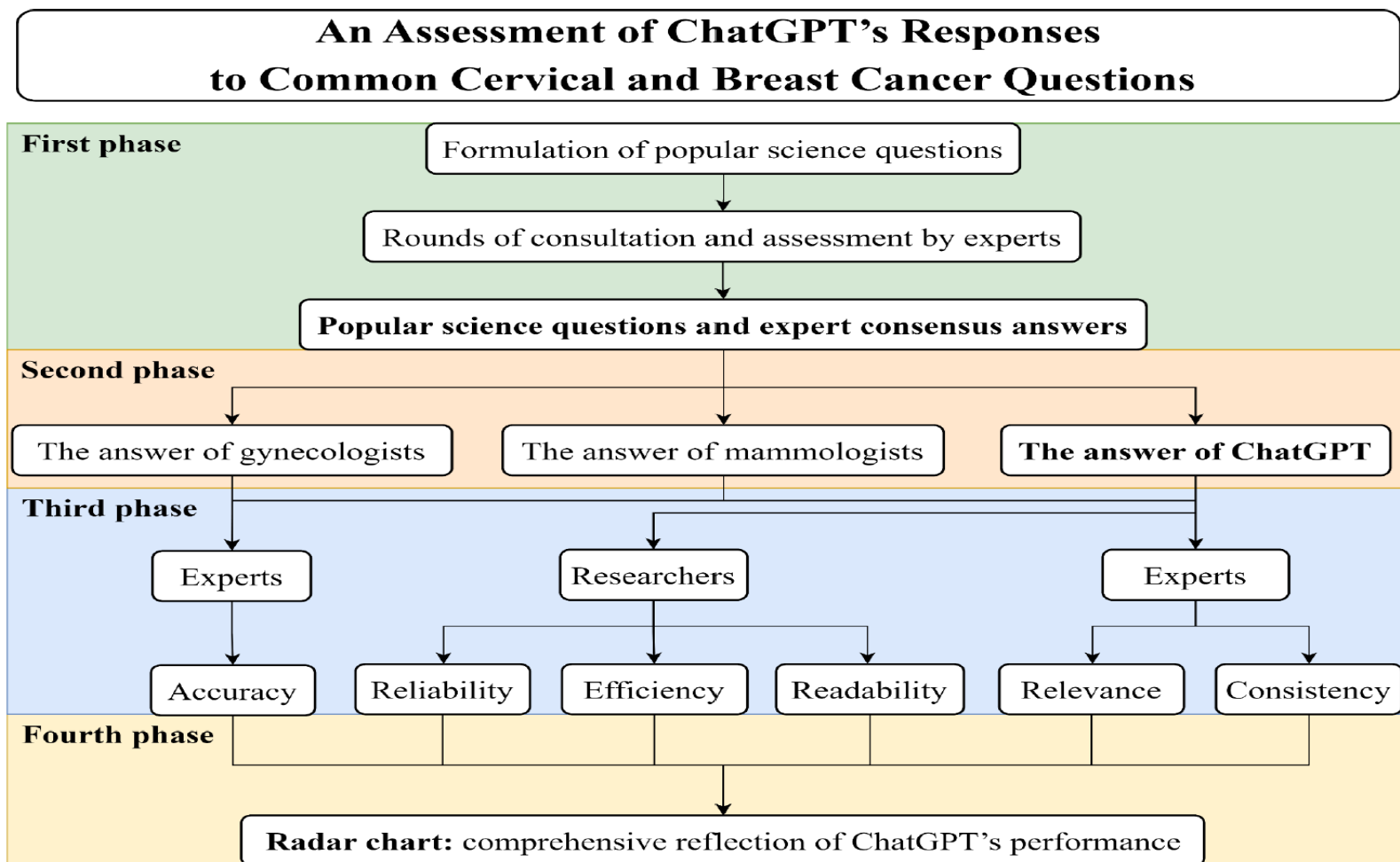
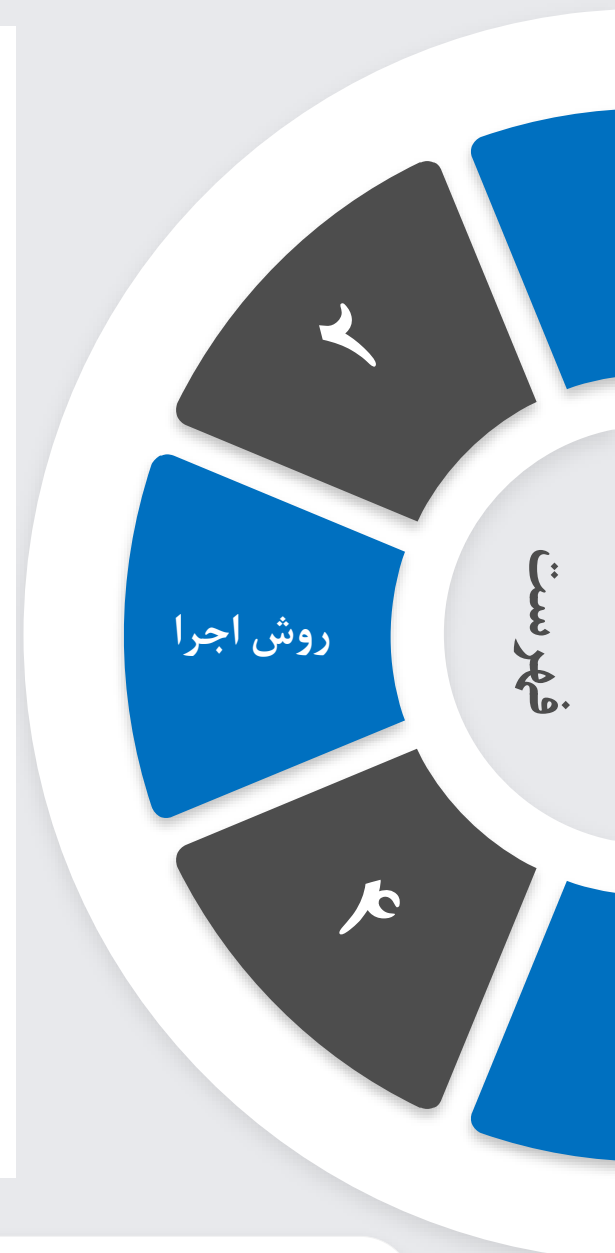
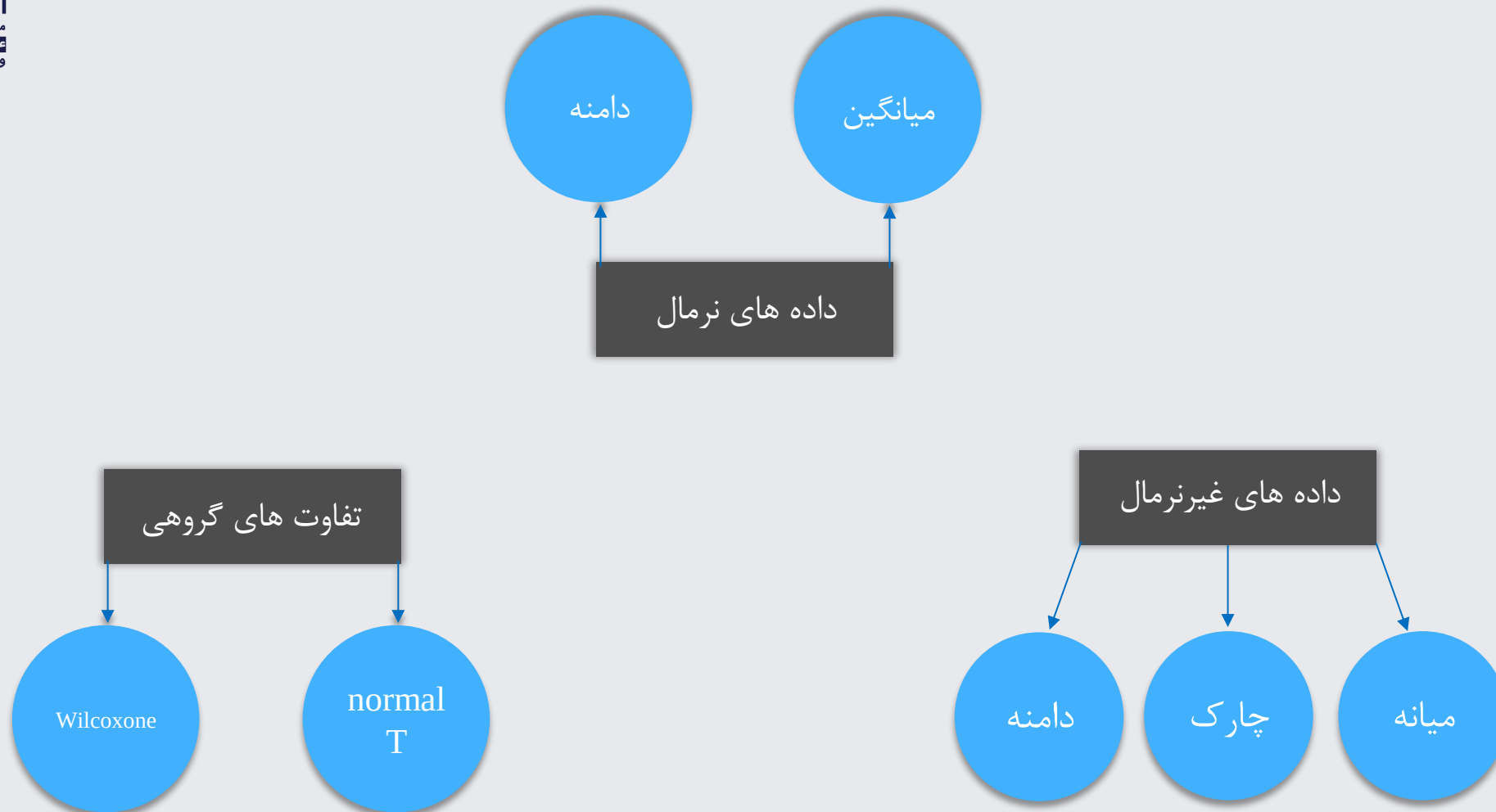


Fig. 1 Flow chart of this study. This study had four main phases. The first phase is to formulate questions and 'reference standard' answers for frequently-asked questions regarding CC and BC. The second phase is to collect the answers from ChatGPT and physicians. The third stage is to evaluate ChatGPT and physicians' answers through six dimensions. The fourth phase is to comprehensively reflect the performance of ChatGPT through radar chart. CC, cervical cancer; BC, breast cancer.



شاخص های ارزیابی ثانویه





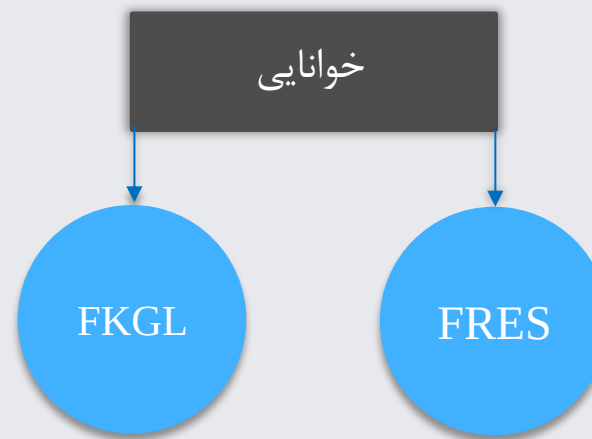
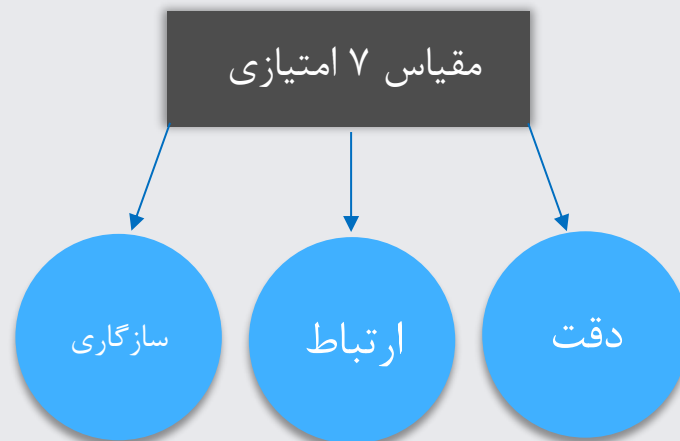


Table 1 Results from Wilcoxon's signed ranks test of accuracy scores

	Median [(P ₂₅ , P ₇₅), (range)]	Difference (P ₂₅ , P ₇₅)	Z	P
Cervical cancer				
Top scoring physician	6.1 [(5.6, 6.4), (4.4-6.8)]	0.8 (0.5, 1.4)	2.657	0.008
ChatGPT	7.0 [(6.6, 7.0), (6.6-7.0)]			
Breast cancer				
Top scoring physician	6.0 [(5.6, 6.4), (3.4-7.0)]	1.0 (0.3, 1.3)	2.527	0.010
ChatGPT	7.0 [(6.9, 7.0), (6.6-7.0)]			
CC and BC				
Top scoring physician	6.0 [(5.7, 6.4), (3.4-7.0)]	0.9 (0.5, 1.4)	3.663	<0.001
ChatGPT	7.0 [(6.7, 7.0), (6.6-7.0)]			

Abbreviations CC, cervical cancer; BC, breast cancer



Table 2 Time, number of words, and readability of answers to CC & BC questions by ChatGPT

	No.	Question	Key assessment answers				Answers to six repeated questions			
			Time	Word count	FREE	FKGL	Time	Word count	FREE	FKGL
Cervical cancer	1	What are the common symptoms of cervical cancer?	9.3	151	40.6	11	7.51	138	54.9	8.7
	2	What are the risk factors for cervical cancer?	20.35	232	39.4	13.6	17.21	194	43	11.2
	3	Why should women get the HPV vaccine?	17.23	243	44.1	13.8	12.63	170	49.2	12.2
	4	How to choose different valent of HPV vaccine?	24	257	49.8	11.6	21.31	238	47.7	11.8
	5	What are the benefits and risks of cervical cancer screening?	8.05	182	30.7	12.8	17.01	248	40.9	11.3
	6	What are the common screening methods for cervical cancer?	12.69	159	53.5	10.2	17.75	209	45.6	11.2
	7	What should I do if my cervical cancer screening result is abnormal?	11.75	159	38.4	11.9	12.91	170	33.5	13.4
	8	Is cervical cancer curable?	14.43	173	35.1	13.1	10.52	118	45.2	11.7
	9	What are the treatment methods for cervical cancer?	19.89	241	40.1	11.2	14.32	178	34.1	12.5
Breast cancer	10	What is the prognosis for cervical cancer?	13.2	158	34.5	13.4	12.36	148	35.4	14.1
	1	What are the common symptoms of breast cancer?	7.58	108	61.5	7.1	7.58	122	70.1	5.9
	2	What are the risk factors for breast cancer?	15.5	238	49.6	11.7	12.52	201	50.2	10.1
	3	How to prevent breast cancer?	13.37	258	52.8	10.5	12.58	251	55.7	9.4
	4	Can breast cancer be inherited?	12.34	178	48.8	12	8.25	113	46	12.4
	5	Why is breast cancer screening important?	13.58	194	51.8	10.9	8.14	126	40.4	33.3
	6	What are the risks of breast cancer screening?	14.2	248	43.6	11.9	11.23	202	29.3	21.7
	7	What are the common screening methods for breast cancer?	9.19	160	55.2	9.5	9.4	162	51.1	10.1
	8	What is the survival rate of breast cancer?	11.18	147	46.6	12.8	8.5	123	45.2	13.4
	9	How to treat breast cancer?	11.9	220	48.1	10.2	11.66	202	39.5	11.4
	10	Can breast cancer recur?	12.39	141	34.4	15.5	7.99	96	45.4	12.3
Mean			13.6	192	44.9	11.7	12.07	170	45.1	12.9
Max			24.0	260	73.7	15.5	21.31	251	70.1	33.3
Min			7.6	108	34.4	4.5	7.51	96	29.3	5.9

Abbreviations: FRES, Flesch Reading Ease Score; FKGL, Flesch-Kincaid Grade Level; HPV, Human papillomavirus



Table 3 Results of consistency test, difference comparison of ChatGPT's answers with six repetitions

Consistency test with six repetitions		A1 to A3		A4 to A6		A1 to A6	
		Count	Median [(P ₂₅ , P ₇₅), (range)]	Count	Median [(P ₂₅ , P ₇₅), (range)]	Count	Median [(P ₂₅ , P ₇₅), (range)]
Cervical Cancer	Differentiation of answering styles						
	Point-by-point response	6	7.0 [(7.0, 7.0), (5.0-7.0)]	4	6.0 [(5.0, 7.0), (3.0-7.0)]	4	6.0 [(5.8, 6.0), (3.0-7.0)]
	Comprehensive response	3	7.0 [(6.0, 7.0), (5.0-7.0)]	4	6.0 [(5.8, 7.0), (5.0-7.0)]	3	5.0 [(5.0, 6.0), (4.0-7.0)]
	Both of them	1	4.0 [(5.0, 5.0), (3.0-5.0)]	2	5.0 [(5.0, 6.0), (3.0-6.0)]	3	5.0 [(3.0, 5.0), (3.0-6.0)]
	Overall consistency score*	10	7.0 [(5.8, 7.0), (4.4-7.0)]	10	6.0 [(5.2, 6.8), (4.1-7.0)]	10	5.1 [(4.6, 6.1), (4.4-6.7)]
Breast cancer	Point-by-point response	6	7.0 [(7.0, 7.0), (5.0-7.0)]	3	7.0 [(6.0, 7.0), (5.0-7.0)]	3	6.0 [(6.0, 6.0), (5.0-7.0)]
	Comprehensive response	4	4.0 [(3.0, 6.0), (3.0-7.0)]	4	5.0 [(3.8, 5.3), (3.0-6.0)]	4	4.0 [(3.0, 5.0), (2.0-5.0)]
	Both of them	0	-	3	5.0 [(5.0, 5.0), (3.0-6.0)]	3	5.0 [(5.0, 5.0), (3.0-7.0)]
	Overall consistency score*	10	6.6 [(4.7, 7.0), (3.3-7.0)]	10	5.4 [(4.4, 6.1), (4.1-7.0)]	10	5.3 [(4.1, 5.7), (3.4, 6.4)]
CC and BC	Point-by-point response	12	7.0 [(7.0, 7.0), (5.0-7.0)]	7	6.0 [(6.0, 7.0), (3.0-7.0)]	7	6.0 [(6.0, 6.0), (3.0-7.0)]
	Comprehensive response	7	6.0 [(4.0, 7.0), (3.0-7.0)]	8	5.0 [(5.0, 6.0), (3.0-7.0)]	7	5.0 [(4.0, 6.0), (3.0-7.0)]
	Both of them	1	4.0 [(5.0, 5.0), (3.0-5.0)]	5	5.5 [(5.0, 6.0), (3.0-6.0)]	6	5.0 [(5.0, 5.0), (3.0-6.0)]
	Overall consistency score*	20	7.0 [(5.6, 7.0), (3.3-7.0)]	20	5.6 [(4.5, 6.5), (4.1-7.0)]	20	5.2 [(4.4, 5.9), (3.4-6.7)]
Difference comparison							
		Median (P ₂₅ , P ₇₅)	Difference (P ₂₅ , P ₇₅)	Z	P		
Cervical Cancer	A1 to A3	7.0 (5.8, 7.0)	0.6 (0.0, 1.5)	1.693	0.090		
	A4 to A6	6.0 (5.2, 6.7)					
Breast cancer	A1 to A3	6.6 (4.7, 7.0)	0.0 (-0.6, 1.8)	1.352	0.180		
	A4 to A6	5.4 (4.4, 6.1)					
CC and BC	A1 to A3	7.0 (5.6, 7.0)	0.2 (0, 1.5)	2.014	0.040		
	A4 to A6	5.6 (4.5, 6.5)					

* Inconsistency test with six repetitions, the overall consistency score is calculated from the average scores of all experts, and the rest scores are calculated from the original scores of all experts

Abbreviations CC, cervical cancer; BC, breast cancer; A1, answer 1; A3, answer 3; A6, answer 6



Table 4 Comparison of the differences in ChatGPT's answer time, word count, and readability between the CC & BC with and without context 1

		Mean (SD)	T	P	Cohen's D (95% CI)
Time	Question	11.9 (2.9)	3.149	0.003	0.996 (0.331, 1.649)
	Question + Con-text	8.9 (3.0)			
Word	Question	167 (45)	3.462	0.001	1.095 (0.422, 1.755)
	Question + Con-text	118 (46)			
FREE	Question	42.1 (10.9)	0.202	0.840	0.064 (-0.556, 0.684)
	Question + Con-text	41.4 (10.3)			
FKGL	Question	12.5 (1.9)	0.098	0.920	0.031 (-0.589, 0.65)
	Question + Con-text	12.5 (1.9)			

Abbreviations FRES, Flesch Reading Ease Score; FKGL, Flesch-Kincaid Grade Level; SD, standard deviation; CI, confidence interval



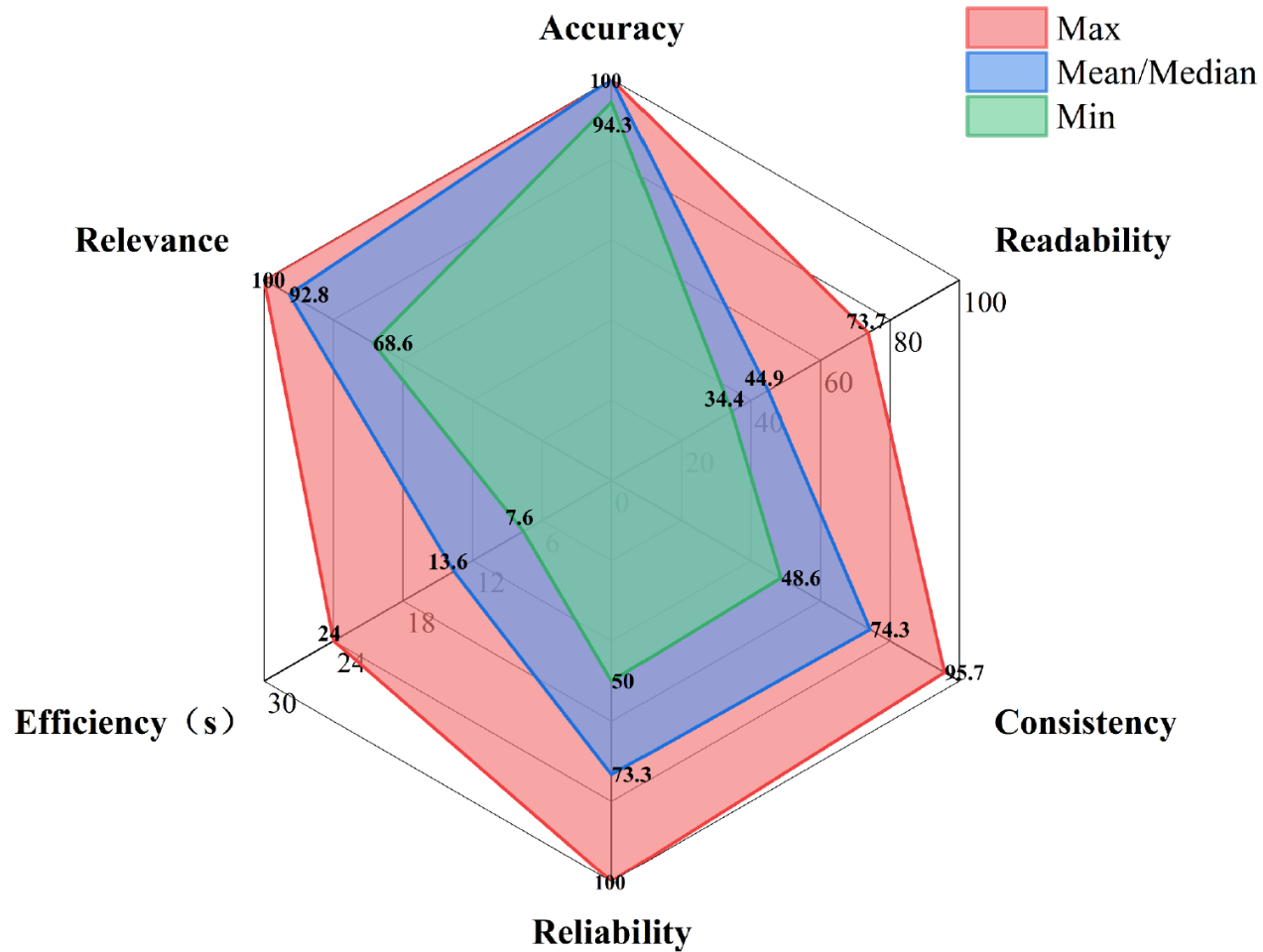
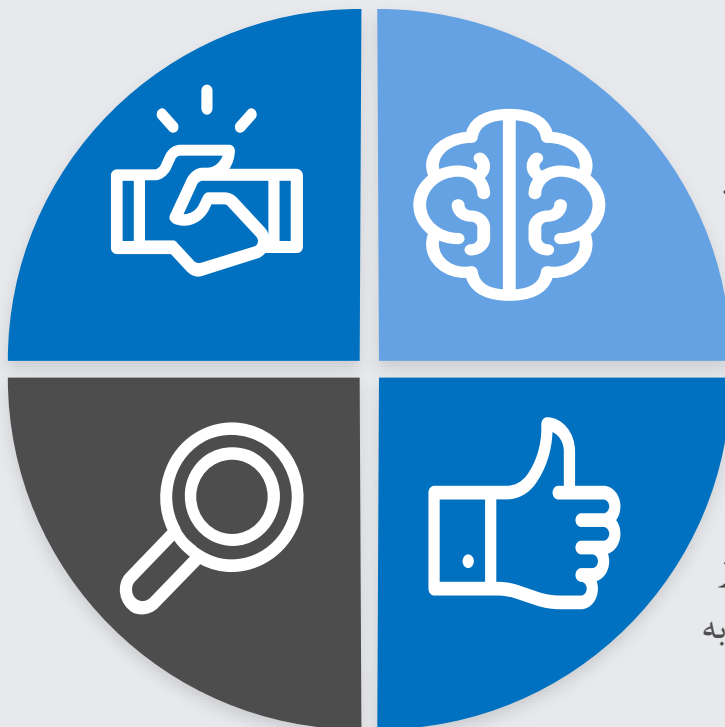


Fig. 2 Radar chart assessing ChatGPT's responses to cervical and breast cancer questions. The performance of ChatGPT was visually assessed with the accuracy, readability, consistency, reliability, efficiency, and relevance through radar images, to reflect the feasibility of applying ChatGPT to the comprehensive prevention and treatment of CC and BC. CC, cervical cancer; BC, breast cancer





پاسخ‌های ChatGPT به سوالات عمومی درباره سرطان‌ها کمی متناقض است. به خصوص در مورد سوالات فردی مانند محدوده سنی برای تجویز واکسن HPV و غربالگری سرطان پستان که ممکن است تا حدی به خاطر عدم **سازگاری** در دستورالعمل‌های سازمان‌های پیشرو باشد.



در بررسی **ارتباط**، ChatGPT نتایج بهتری در پاسخ به سوالات در زمینه‌های مختلف نسبت به بدون زمینه‌های اضافی ارائه می‌دهد.

خوانایی کم و طولانی که حداقل به سطح تحصیلات دبیرستانی نیاز دارد. طبق دستورالعمل‌ها انتقال اطلاعات پیچیده با جملات کوتاه موثرتر است.

تحلیل **قابلیت اطمینان** نشان داد که نیمی از منابع استفاده شده توسط ChatGPT مربوط به قبل سپتامبر ۲۰۲۱ بوده و باید به روز شوند. همچنین اتکا به استاندارد‌های آمریکایی در پاسخ به سوالات انگلیسی ممکن است دقت و ارتباط آن را در دیگر مناطق محدود کند.



محدودیت ها

- ✓ **دقت** پاسخ ها تنها یک بار ارزیابی شد که این احتمال خطاهای تصادفی را به وجود می آورد.
- ✓ **معیار ارزیابی** عمدتاً به ارزیابی های کارشناسان متکی بود که با وجود توافق خوب بین کارشناسان، هنوز دارای برخی درجه های ذهنی و پتانسیل تغییر الگو است.
- ✓ این مطالعه تنها بر روی سوالات به **زبان** انگلیسی تمرکز کرده و ارزیابی عملکرد ChatGPT در زبان های دیگر را محدود می کند.
- ✓ تعداد محدود سوالات درباره CC و BC با **تفاوت های ملی/منطقه ای** که نیاز به تحقیقات بیشتری برای بررسی تغییرات در مناطق مختلف دارد.



دیدگاه من

۱. پرسش سوال ها با زبان های دیگر
۲. افزایش اعضا و تنوع افراد استاندارد مرجع
۳. استفاده از GPT-4



با تشکر از توجه شما



Breast cancer

No.	Question
1	What are the common symptoms of breast cancer?

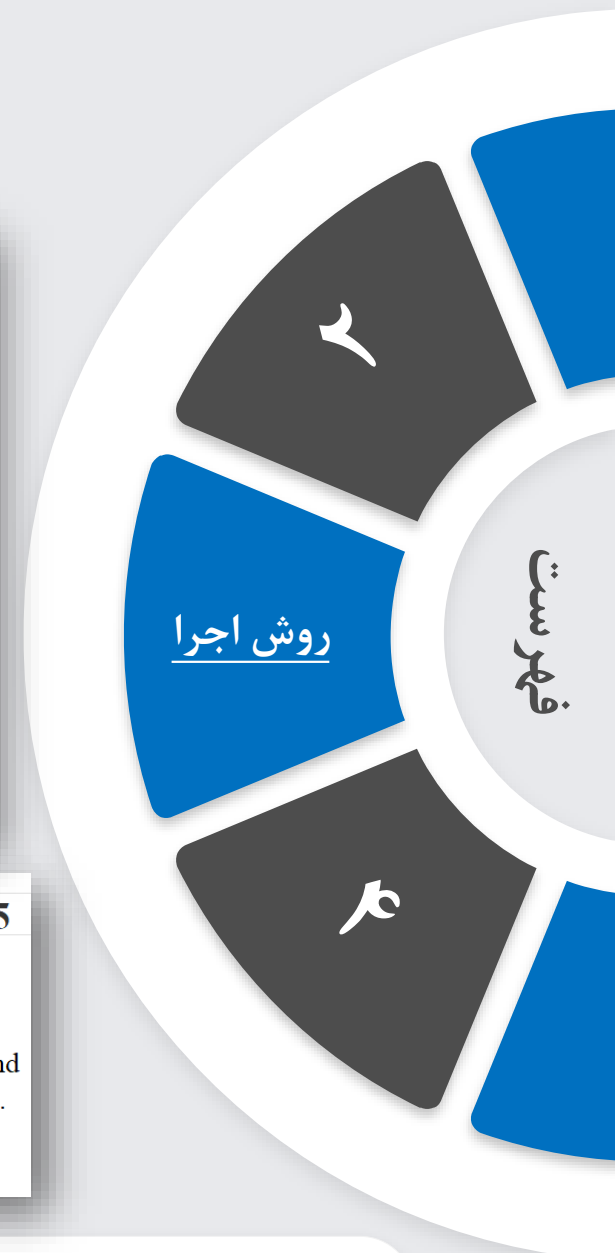
Expert consensus

Early-stage breast cancer usually does not have signs or symptoms, making it hard to detect. Symptoms usually manifest in intermediate to advanced stages of breast cancer. It is important to recognize that symptoms of breast cancer can vary among individuals. Certain warning signs of breast cancer include:

- (1) Development of a new lump in the breast or underarm (armpit).
- (2) Any changes in the size or shape of the breast.
- (3) Thickening or swelling of a specific area in the breast.
- (4) Pain experienced in any region of the breast.
- (5) Irritation or dimpling of the skin on the breast.
- (6) Redness or flaky skin around the nipple or breast area.
- (7) Retraction or inward pulling of the nipple or pain in the nipple area.
- (8) Discharge from the nipple that is not breast milk, including blood.

It is important to note that these symptoms can also be associated with conditions other than cancer. If you experience any signs or symptoms that concern you, it is advisable to promptly consult your doctor.

Mammologists 1	Mammologists 2	Mammologists 3	Mammologists 4	Mammologists 5
Less active lump in the breast, which is painless and grows rapidly in the short term, There may appear some skin changes in the advanced or terminal stage (Redness skin, orange peel skin, edema, etc.).	Painless lump. Changes in breast skin. Nipple discharge.	A painless lump. Nipple discharge.	Lump. Orange peel skin. Pain.	Pain in breast. Nipple discharge. Pulling in of the nipple. Breast skin depression and lymph node enlargement.



Questions

Appendix A: The way of asking questions for ChatGPT.

(1) The original question

Question 1

—What are the common symptoms of cervical cancer?

—What is the basis for your answer above? Please include a link or reference and citation time.

(2) Add contextual questions

Context 1: I am a relatively healthy woman. Please balance readability and conciseness while ensuring accuracy, so that I can easily understand and accept the answers. What are the common symptoms of cervical cancer?

Context 2: I am from China, and I am now in China. What are the common symptoms of cervical cancer? Please answer in English.

Cervical cancer

Access Date: April 26, 2023

Question

1 What are the common symptoms of cervical cancer?

ChatGPT Answer

Cervical cancer often does not cause symptoms in its early stages. However, as the cancer progresses, the following symptoms may occur:

1. Abnormal vaginal bleeding, such as bleeding after sexual intercourse, between periods or after menopause.
2. Pain during sexual intercourse.
3. Unusual vaginal discharge that may be heavy, watery, or have a foul odor.
4. Pelvic pain or pain in the lower back.
5. Urinary problems, such as pain during urination or blood in the urine.
6. Fatigue, weight loss, and loss of appetite in advanced stages of the disease.

It's important to note that these symptoms can also be caused by other conditions, and having these symptoms does not necessarily mean that you have cervical cancer. If you experience any of these symptoms, it's important to talk to your healthcare provider to determine the underlying cause and get appropriate treatment. Additionally, getting regular cervical cancer screenings can help detect the disease early, before symptoms occur.

Reference source

My answer is based on my general knowledge as a language model trained on a large corpus of texts, including medical and scientific literature, as well as reputable health sources.

However, to provide a more specific reference, I can cite the American Cancer Society's website on cervical cancer, which describes the symptoms of the disease. The source can be found at the following link: <https://www.cancer.org/cancer/cervical-cancer/detection-diagnosis-staging/signs-and-symptoms.html>

The information on the website was accessed on September 2021, which was my knowledge cutoff point.

