中华医学会系列杂志

ISSN 1007-5232
CN 32-1463/R

中华消化内镜杂志®

ZHONGHUA XIAOHUA NEIJING ZAZHI

2023年11月 第40卷 第11期

CHINESE JOURNAL OF DIGESTIVE ENDOSCOPY

Volume 40 Number 11
November 2023

中华医学会
CHINESE MEDICAL ASSOCIATION

ISSN 1007-5232
11>

·论著·

消化医护人员与新型聊天机器人在消化内镜癌症筛查相关知识的调查表现

刘亚玲[1]　吕勇[1]　康晨曦[1]　王向平[1]　李静[1]　王陵[2]　王海英[1]　聂勇战[1]　吴开春[1]　潘阳林[1]

[1]空军军医大学西京消化病医院 国家消化系统疾病临床医学研究中心 消化系肿瘤整合防治全国重点实验室，西安 710032；[2]空军军医大学统计学教研室，西安 710032

通信作者：潘阳林，Email：yanglinpan@hotmail.com

【摘要】 目的　探讨并比较消化医护人员以及新型聊天机器人（ChatGPT 和新必应）对消化内镜癌症筛查相关知识的知晓情况。**方法**　问卷共设置 23 个消化内镜癌症筛查相关问题，涉及消化道癌症的筛查年龄、高危因素、随访时间和消化内镜的优势和风险等。邀请消化医护人员通过电子问卷形式回答问题，应用 ChatGPT 和新必应对每个问题分别进行 10 轮回答。以所有问题答案的正确率作为主要研究终点。采用方差分析比较消化医护人员与新型聊天机器人回答消化内镜筛查癌症相关知识的正确率，采用单因素和多因素线性回归模型分析影响消化医护人员答题正确率的因素。**结果**　共 76 名消化医护人员（规培医师 21 名、消化专科护士 28 名、消化专科医师 27 名）答题结果被纳入分析。规培医师、消化专科护士、消化专科医师、ChatGPT 和新必应的回答正确率分别为 36.4%± 10.9%、34.5%±10.2%、52.2%±12.6%、46.3%±9.8% 和 67.1%±9.3%，差异有统计学意义（F=22.6，P<0.001）。新必应回答正确率最高（P<0.001）；ChatGPT 正确率和消化专科医师相当（LSD-t=−1.398，P=0.166），均高于消化专科护士（LSD-t=2.956，P=0.004；LSD-t=5.955，P<0.001）和规培医师（LSD-t=2.402，P=0.018；LSD-t=4.951，P<0.001）；消化专科护士和规培医师正确率相当（LSD-t=−0.574，P=0.567）。与新必应相比，消化专科医师回答消化道癌症筛查不良事件、肠化生随访、结肠癌高危因素和筛查方法等问题的正确率低（P<0.05），回答内镜不良事件和食管癌筛查方法问题的正确率高（P<0.05）。多因素线性回归模型分析显示消化专科医师（β=11.7，t=3.054，P=0.003）和问卷回答时间（≥7.6 min）（β=7.8，t=2.894，P=0.005）是消化医护人员回答正确率的独立影响因素。**结论**　与消化医护人员相比，新型聊天机器人——新必应回答癌症筛查相关问题的更准确，但在内镜不良事件问题和食管癌筛查方法问题回答方面表现不佳。

【关键词】 内窥镜检查，消化系统；　食管肿瘤；　胃肿瘤；　结直肠肿瘤；　筛查；　机器学习

基金项目：国家重点研发计划（2022YFC2505100，2017YFC0908300）

An investigation of digestive healthcare workers and new chatbots on knowledge of endoscopic screening of gastrointestinal cancers

Liu Yaling[1], Lyu Yong[1], Kang Chenxi[1], Wang Xiangping[1], Li Jing[1], Wang Ling[2], Wang Haiying[1], Nie Yongzhan[1], Wu Kaichun[1], Pan Yanglin[1]

[1]Xijing Hospital of Digestive Diseases, Air Force Medical University, National Clinical Research Center for Digestive Diseases, State Key Laboratory of Holistic Integrative Management of Digestive Cancers, Xi'an 710032, China; [2]Department of Statistics, Air Force Medical University, Xi'an 710032, China

Corresponding author: Pan Yanglin, Email: yanglinpan@hotmail.com

DOI：10.3760/cma.j.cn321463-20230424-00087

收稿日期　2023-04-24　　**本文编辑**　许文立　唐涌进

引用本文：刘亚玲，吕勇，康晨曦，等. 消化医护人员与新型聊天机器人在消化内镜癌症筛查相关知识的调查表现[J]. 中华消化内镜杂志，2023，40(11)：892-899. DOI：10.3760/cma.j.cn321463-20230424-00087.

www.zhxhnjzz.com

【Abstract】 **Objective** To compare the knowledge of endoscopic screening of gastrointestinal cancers between digestive healthcare workers and new chatbots (chatGPT and new Bing). **Methods** A test with twenty-three questions of endoscopic screening of gastrointestinal cancers was conducted, focusing on the appropriate age of screening, high-risk factors, the follow-up time, and the advantages and risks of digestive endoscopy. Digestive healthcare workers were invited to complete the test through electronic questionnaires. New Bing and chatGPT were used to answer each question for 10 rounds. The primary endpoint was the correct rate of all answers. The answer accuracy between digestive healthcare workers and new chatbots were compared using variance analysis, and the factors that affected the accuracy of the answers in digestive healthcare workers were explored using univariate and multivariable liner regression analysis. **Results** The results of the test completed by 76 digestive healthcare workers (21 residents, 28 digestive nurses, and 27 digestive doctors) were analyzed. The accuracies were 36.4%±10.9%, 34.5%±10.2%, 52.2%± 12.6%, 46.3%±9.8% and 67.1%±9.3% in residents, digestive nurses, digestive doctors, chatGPT, and new Bing, respectively, with significant difference (F=22.6, P<0.001). The accuracy was highest in new Bing (P<0.001). The accuracy was comparable between chatGPT and digestive doctors (LSD-t=−1.398, P=0.166), and both higher than that of digestive nurses (LSD-t=2.956, P=0.004; LSD-t=5.955, P<0.001) and residents (LSD-t=2.402, P=0.018; LSD-t=4.951, P<0.001). Furthermore, the accuracy was comparable between digestive nurses and residents (LSD-t=−0.574, P=0.567). Compared with new Bing, digestive doctors had lower accuracy in answering questions related to adverse events of screening, follow-up recommendation of intestinal metaplasia, high risk factors and screening methods for colon cancer (P<0.05), but higher accuracy in answering questions related to endoscopic adverse events and screening methods for esophageal cancer (P<0.05). Multivariable liner regression analysis showed that being digestive doctors (β=11.7, t=3.054, P=0.003) and questionnaire response time (≥7.6 min) (β =7.8, t=2.894, P=0.005) were independent factors for the answer accuracy of digestive healthcare workers. **Conclusion** Compared with digestive healthcare workers, New chatbots—new Bing has higher accuracy in answering gastrointestinal cancer screening-related questions, but performs poorly in answering questions such as adverse events of endoscopy and screening methods for esophageal cancer.

【Key words】 Endoscopy, digestive system; Esophageal neoplasms; Stomach neoplasms; Colorectal neoplasms; Screening; Machine learning

Fund program: National Key Research and Development Program (2022YFC2505100, 2017YFC0908300)

消化道癌症是危害人类健康的重大疾病，在我国发病率和死亡率排前十位的癌症中，消化道癌（食管癌、胃癌、结直肠癌）约占 1/2，其防治任务任重道远。然而，消化道癌筛查有利于早期诊断，从而提高治愈率和生存率。近五年有十余部消化道癌筛查的指南或共识意见颁布[1-11]，为消化道癌筛查提供了规范意见。

消化内镜是食管癌、胃癌和结直肠癌筛查的主要方法之一，值得大力推广。但由于筛查知识缺乏、参与过程繁琐及担忧风险和费用等问题，我国参与消化内镜早期癌筛查的人群比例较低。以结肠癌筛查为例，发达国家中老年人参与率为 20%~60%[12]，而我国大部分地区结肠癌筛查参与率不足 5%[13]。医护人员的宣传、解释和引导，可有效提升人们参与癌症筛查的意愿。然而，消化医护人员自身对消化道癌及消化内镜相关筛查知识的掌握情况尚需要进一步评价。目前互联网已成为查找癌症及相关筛查知识的重要途径，但传统搜索工具（如百度、必应等）检索到的癌症筛查相关信息混杂，非专业人士难以辨别正确性。新型聊天机器人如 ChatGPT 和新必应被认为在回答一些问题的准确性上有质的提升。ChatGPT 的应用程序编程接口现面向全世界用户开放，同时新必应邀请受试人群的规模显著扩大，上述两种新型聊天机器人目前在国内均可获得使用权限。本研究也综合探索这两种工具在中国消化道癌症筛查相关问题中的表现。

资料与方法

一、消化内镜癌症筛查问卷

参考 2020—2022 年颁布的 11 部中国消化道癌筛查相关指南设置问卷[1-11]，共设置 23 个问题，其中单选 14 题，多选 9 题。分别包括食管癌、胃癌、结直肠癌的筛查年龄范围、筛查方法、高危人群、内镜诊断或治疗后随访等，还涉及适宜筛查的癌症种类、消化内镜筛查的优势和风险等。以电子问卷的形式（金数据，www.jinshuju.net）发放。

www.zhxhnjzz.com

二、医护人员

本研究邀请空军军医大学西京消化病医院的规培医师、消化专科护士以及消化专科医师参与问卷调查，各邀请人数为20人以上，要求参与人员不与他人讨论问卷答案，但不限制通过互联网或专业书籍等寻找答案。收集所有消化医护人员的性别、年龄、学历、从业年限等资料；对于消化专科医师，还收集完成结肠息肉切除手术量、年完成内镜黏膜下剥离术（endoscopic submucosal dissection，ESD）手术量等信息。

为保证消化医护人员回答调查问卷的质量，规定每个问题平均回答时间至少5 s，基本信息填写20~30 s，因此每个参与人员答题时间不少于3 min，对答题时间<3 min的参与人员予以剔除。

三、新型聊天机器人

分别在ChatGPT和新必应的聊天对话框中输入23个问题，获得问题答案。每一个问题完成10轮作答。当回答方式为描述性（未直接给出确定答案）时，由两名消化道癌筛查专家判别答案的准确性，若两名专家意见不一致，则由第3名专家商讨后决定；当出现模棱两可的回答时，判别为错误。由于两种新型聊天机器人每一轮回答的答案可能存在差异，为了测试答案分布情况，选择“食管早癌ESD术后随访”问题（问题11）进行测试，完成100轮提问。

四、研究终点

本研究的主要研究终点为消化医护人员以及新型聊天机器人回答所有23个内镜筛查相关问题的平均正确率。比较各类别消化医护人员与新型聊天机器人的平均正确率；分析影响消化医护人员平均正确率的因素；对于高错误率（>70%）问题，以消化专科医师和新必应分别作为消化医护人员和新型聊天机器人的代表，分析两者在23个高错误率问题中的错误答案集中分布情况。

五、统计学方法

利用统计学软件PASS 15计算样本量，SPSS 26进行统计分析。为比较消化专科医师和新必应所有23个问题的平均正确率，计算两样本t检验所需样本量，根据既往研究[14-15]，初步估计消化专科医师和新必应的回答正确率分别为50%和75%，标准差均为20%，其比例为2∶1，设定α为0.05，β为0.20，计算出样本量分别为18和9。计量资料满足正态分布或近似正态分布时采用$\bar{x}\pm s$表示，否则采用M（IQR）表示。采用ANOVA检验分析比较各类别消化医护人员、新型机器人的平均正确率，若差异存在统计学意义，则采用LSD-t检验进行事后组间多重比较。计数资料以频数（%）表示，组间比较采用卡方检验或Fisher确切概率法。使用线性回归分析医护人员回答正确率的影响因素，人员类别中以规培医师为对照设置哑变量，单因素线性回归分析中P<0.1的变量纳入多因素线性回归分析，结果用β（95%CI）表示。使用百分条图描述单一答案的分布情况。P<0.05为差异有统计学意义。

结　果

一、参与问卷调查医护人员的基本情况

本研究共入选77名消化医护人员参与电子问卷调查，排除作答敷衍1名（约1 min完成），共76名纳入分析，其中规培医师、消化专科护士和消化专科医师分别为21名、28名和27名（表1）。纳入分析的消化医护人员年龄为35（11）岁，女性占69.7%（53/76）。消化专科医师年龄为39（12）岁，从业时间11（14）年，开展消化内镜诊疗工作8（12）年，年完成结肠息肉切除100（395）例，年完成ESD 0（30）例（有ESD经验者10名）。消化医护人员问卷中位作答时间4~9 min，11.8%（9/76）在作答过程中寻求了互联网、专业书籍等帮助。

表1　参与调查消化医护人员的基线资料

项目	规培医师（n=21）	消化专科护士（n=28）	消化专科医师（n=27）
年龄[岁，M（IQR）]	25（3）	37（3）	39（12）
女性[名（%）]	16（76.2）	26（92.9）	11（40.7）
学历[名（%）]			
本科及以下	10（47.6）	28（100.0）	3（11.1）
本科以上学历	11（52.4）	0（0.0）	24（88.9）
执业工作年限[年，M（IQR）]	0（2）	14（4）	11（14）
开展消化内镜时间[年，M（IQR）]	-	-	8（12）
年结肠息肉切除量[例，M（IQR）]	-	-	100（395）
年ESD数量[例，M（IQR）]	-	-	0（30）
作答时间[min，M（IQR）]	6（3）	9（5）	8（5）
寻求帮助[名（%）]	4（19.0）	3（10.7）	2（7.4）

注：“-”指不具备内镜操作资格，相应资料不做记录；ESD指内镜黏膜下剥离术

www.zhxhnjzz.com

二、医护人员及新型聊天机器人作答情况

如表2显示，规培医师、消化专科护士、消化专科医师、ChatGPT和新必应的平均正确率分别为36.4%±10.9%、34.5%±10.2%、52.2%±12.6%、46.3%±9.8%和67.1%±9.3%，差异有统计学意义（F=22.6，P<0.001）。LSD-t检验事后多重比较中，新必应的正确率最高（P<0.001）；ChatGPT与消化专科医师的正确率相当（LSD-t=−1.398，P=0.166），均高于消

表2　医护人员及新型聊天机器人回答消化道癌筛查相关问题的准确性

23个消化内镜筛查相关问题及答案	规培医师（n=21）	消化专科护士（n=28）	消化专科医师（n=27）	ChatGPT（n=10）	新必应（n=10）	P值[a]
1.在中国发病最高的前十大癌症中，您认为适合早期筛查的癌症有哪些（可多选）？**A 肺癌；B 结直肠癌；C 胃癌；D 乳腺癌；E 肝癌；F 食管癌**；G 甲状腺癌；H 胰腺癌；I 前列腺癌；**J 宫颈癌**	23.8%	17.9%	22.2%	0.0%	100.0%	<0.001
2.如果参与食管、胃、结直肠癌的筛查，可能带来的不良后果是什么（可多选）？**A 过度诊断问题；B 筛查带来的精神紧张、焦虑；C 漏诊问题；D 检查和随访的费用增加；E 内镜检查的潜在风险**	0.0%	21.4%	0.0%	60.0%	90.0%	<0.001
3.在中国，开展消化道癌筛查时，选择内镜筛查的优点有哪些（可多选）？**A 准确性最高；B 如发现早期病变，可术中完成治疗**；C 舒适度最高；D 安全无风险；E 适用于各年龄段人群	33.3%	10.7%	25.9%	80.0%	30.0%	1.000
4.在消化道癌筛查时，选择胃肠镜筛查最常见的不良反应是什么？A 出血；B 穿孔；C 病变漏诊；D 误吸风险；**E 术后腹部不适感**	66.7%	57.1%	44.4%	0.0%	0.0%	0.015
5.在中国，普通人群参与食管癌筛查，起始的年龄应该是什么？A 30岁；**B 40~50岁**；C 60~70岁；D 目前尚不确定	76.2%	82.1%	88.9%	100.0%	50.0%	0.021
6.在中国，普通人群参与食管癌筛查，最大年龄（即年龄上限）应该是多少？A 55岁；B 65岁；**C 75岁**；D 85岁；**E 目前尚不确定**	71.4%	53.6%	77.8%	0.0%	90.0%	0.647
7.在中国开展食管癌筛查，应选择哪些方法进行筛查（可多选）？A 肿瘤标志物；B 上消化道钡餐；**C 胃镜**；D ^{13}C呼气试验；**E 新型细胞收集器**	19.0%	7.1%	44.4%	0.0%	10.0%	0.065
8.中国人食管癌相关的高危因素有哪些（可多选）？**A 吸烟；B 饮酒；C 食管癌高发区长期居住；D 热烫食物；E 食管癌家族史**	47.6%	57.1%	63.0%	100.0%	70.0%	1.000
9.在中国，参与食管癌筛查的人群，如果胃镜发现食管黏膜是完全正常的，那么下一次复查胃镜的时间应该是什么时候？A 1年；B 3年；**C 5~10年**；D 10年	9.5%	7.1%	40.7%	30.0%	100.0%	0.002
10.在内镜检查中，如果发现存在巴雷特食管（确定无异型增生），那么下一次复查内镜的时间应该选择什么时候？A 1年；**B 3~5年**；C 5~10年；D 10年	4.8%	25.0%	37.0%	40.0%	70.0%	0.136
11.对于内镜发现的食管高级别上皮内瘤变，ESD完整切除后无食管狭窄者，下一次胃镜随访的时间是什么时候？A 1个月；**B 3~6个月**；C 6~12个月；D 1~2年；E 2~3年	57.1%	42.9%	77.8%	60.0%	40.0%	0.049
12.在中国，普通人群参与胃癌筛查，起始的年龄应该是什么？A 30岁；**B 40-50岁**；C 60~70岁；D 目前尚不确定	47.6%	75.0%	92.6%	100.0%	60.0%	0.035
13.在中国，普通人群参与胃癌筛查，最大年龄（即年龄上限）应该是多少？A 55岁；B 65岁；**C 75岁**；D 85岁；**E 目前尚不确定**	71.4%	42.9%	74.1%	10.0%	100.0%	0.155
14.在中国开展胃癌筛查，应选择哪些方法进行筛查（可多选）？A 肿瘤标志物如CEA、CA19-9或CA125等；B 腹部B超；**C 胃镜；D ^{13}C呼气试验；E PG检测**	14.3%	3.6%	33.3%	0.0%	100.0%	<0.001
15.中国人胃癌相关的高危因素有哪些（可多选）？**A 吸烟；B 饮酒；C *HP*感染；D 烟熏食物；E 胃癌家族史**	42.9%	75.0%	81.5%	100.0%	70.0%	0.655
16.在中国，参与胃癌筛查的人群，如果胃镜发现胃黏膜是完全正常的，那么下一次复查胃镜的时间应该是什么时候？A 1年；B 3~5年；C 5~10年；D 10年；**E 根据不同危险度分级选择1~5年复查**	28.6%	28.6%	37.0%	100.0%	90.0%	0.008
17.在内镜检查中，如果发现存在慢性萎缩性胃炎（C1）伴肠化生，那么下一次复查内镜的时间应该是什么时候？A 1年；**B 3年**；C 5~10年；D 10年	14.3%	3.6%	18.5%	30.0%	30.0%	0.655
18.对于内镜发现的胃内高级别上皮内瘤变，ESD完整切除后，下一次胃镜随访的时间是什么时候？A 1个月；**B 3~6个月**；C 6~12个月；D 1~2年；E 2~3年	57.1%	46.4%	77.8%	60.0%	90.0%	0.647
19.在中国，普通人群参与结肠癌筛查，起始的年龄应该是什么？A 30岁；**B 40~50岁**；C 60~70岁；D 目前尚不确定	61.9%	71.4%	92.6%	100.0%	30.0%	<0.001
20.在中国，普通人群参与结直肠癌筛查，最大年龄（即年龄上限）应该是多少？A 55岁；B 65岁；**C 75岁**；D 85岁；**E 目前尚不确定**	66.7%	46.4%	74.1%	20.0%	100.0%	0.155
21.在中国开展结直肠癌筛查，应选择哪些方法进行筛查（可多选）？A 肿瘤标志物如CEA、CA19-9或CA125等；**B 结肠镜；C FIT**；D 肠道菌群分析；E 钡剂灌肠	0.0%	0.0%	11.1%	10.0%	50.0%	0.021
22.中国人结直肠癌相关的高危因素有哪些（可多选）？**A 吸烟；B 饮酒；C 肥胖**；D 便秘；**E 结肠癌家族史**	14.3%	0.0%	14.8%	20.0%	70.0%	0.003
23.在中国，参与结肠癌筛查的人群，如果高质量的结肠镜检查发现结肠黏膜是完全正常的，那么下一次复查结肠镜的时间应该是什么时候？A 1年；B 3年；**C 5~10年**；D 无需再查	9.5%	21.4%	70.4%	50.0%	100.0%	0.079

注：加粗的答案为每个题目正确答案；问题1跨越多癌学科、有10项选择且各指南观点不同，故正确项≥3项即计为答案正确（G、H、I为错误选项）；问题6、13、20因各指南有所差异，故B或者E均正确；问题14因各指南有所差异，故C、C+D、C+E、C+D+E四种均为正确；ESD指内镜黏膜下剥离术；PG指胃蛋白酶原；*HP*指幽门螺杆菌；FIT指免疫化学法粪便隐血试验；[a]新必应与消化专科医师比较

www.zhxhnjzz.com

化专科护士(LSD-t=2.956, P=0.004; LSD-t=5.955, P<0.001)和规培医师(LSD-t=2.402, P=0.018; LSD-t=4.951, P<0.001);消化专科护士和规培医师水平相当(LSD-t=−0.574, P=0.567)。

鉴于新必应与消化专科医师回答问题正确率分别代表聊天机器人和消化医护人员的最高水平,对两者回答单一问题的正确率进一步比较。在23个筛查相关问题中,新必应对问题1、2、9、14、16、21、22的回答正确率高于消化专科医师(P<0.05),对问题4、5、11、12、19回答正确率低于消化专科医师(P<0.05),其余问题两者回答正确率相当(表2)。

采用线性回归模型对影响医护人员回答正确率的因素进行分析,单因素分析发现,男性、消化专科护士、消化专科医师和本科以上学历是消化医护人员回答正确率的影响因素(P<0.05)。多因素分析显示,消化专科医师和作答时间是消化医护人员回答正确率的独立影响因素(P<0.05)。具体见表3。

三、高错误率问题分析

在本研究中消化专科医师和新必应在回答34.8%(8/23)的消化道癌筛查相关问题上表现不佳,错误率高达70%以上(表4)。有6个问题消化专科医师回答不佳,主要集中在癌症筛查不良事件、胃黏膜肠化生随访决策和结肠癌高危因素等问题上;新必应在2个问题上回答不佳,80.0%(8/10)答案误认为出血或穿孔是内镜的最常见不良事件,90.0%(9/10)答案误认为上消化道钡餐检查适合用于食管癌筛查。

四、答案分布一致性分析

为了测试新型聊天机器人在同一问题回答上的一致情况,ChatGPT和新必应分别回答100次问题11,答案分布情况如图1所示,ChatGPT和新必应在100轮测试中主要分布略相似,ChatGPT主要分布于B(53.0%)和C(39.0%)选项,新必应主要分布于C(54.0%)和B(38.0%)选项,此外两种新型聊天机器人有可能出现拒绝回答或者不确定的情况。

表3　影响消化医护人员回答问题正确率因素的线性回归分析

项目	单因素分析			多因素分析		
	β值(95%CI)	t值	P值	β值(95%CI)	t值	P值
年龄>40岁	5.9(−1.9~44.2)	1.479	0.139			
性别男性	6.8(0.1~13.5)	2.029	0.046	−0.6(−6.9~5.6)	−0.200	0.842
人员类别						
规培医师[a]	0			0		
消化专科护士	−10.7(−16.7~4.6)	−3.494	0.001	−2.6(−10.5~5.2)	−0.663	0.509
消化专科医师	16.8(11.4~22.1)	6.241	<0.001	11.7(4.1~19.3)	3.054	0.003
本科以上学历	12.9(7.3~18.5)	4.588	<0.001	5.7(−2.1~13.6)	1.468	0.147
执业年限≥6年	−0.6(−7.3~6.2)	−0.166	0.868			
作答时间≥7.6 min	6.0(−0. 2~12.1)	1.918	0.059	7.8(2.4~13.1)	2.894	0.005
作答寻求帮助	−4.0(−15.7~7.7)	−0.680	0.498			

注:[a]对照组

表4　高错误率消化内镜癌症筛查问题的错误原因

问题	主要错误原因	
	消化专科医师(n=27)	新必应(n=10)
1.适合筛查的癌症种类	77.8%(21名)误认为甲状腺癌适合癌症筛查	−
2.癌症筛查相关风险	96.3%(26名)选择筛查不良事件≤3条,风险认识不全面	−
3.内镜筛查的优点	59.3%(16名)误认为各年龄段人群均适合内镜筛查	−
4.内镜的不良事件	−	80.0%(8名)误认为出血或穿孔是内镜的最常见不良事件
7.食管癌筛查方法	−	90.0%(9名)误认为上消化道钡餐检查适合用于食管癌筛查
17.胃黏膜肠化生的复查	81.5%(22名)误认为应该1年后随访	−
21.结直肠癌筛查方法	77.8%(21名)误认为传统肿瘤标志物可用于结肠癌筛查	−
22.结直肠癌高危因素	81.5%(22名)误认为便秘是结直肠癌的高危因素	−

注:高错误率问题指错误率>70%;“−”指该问题在消化专科医师或者新必应中不是高错误率问题

www.zhxhnjzz.com

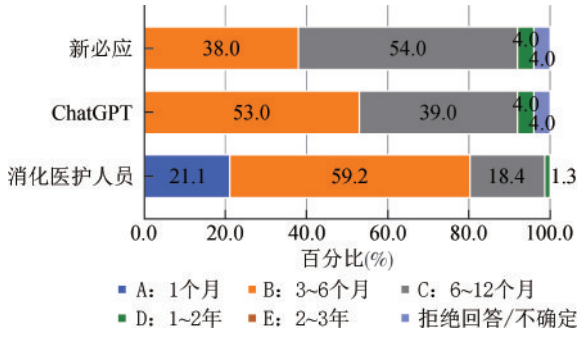


图1　消化医护人员、ChatGPT和新必应回答问题11的答案分布

讨　论

消化内镜是消化道癌症筛查的主要方法之一，与幽门螺杆菌检测、粪便潜血检测、食管新型细胞收集器等筛查方法一样，其有相应的优缺点和适应范围，需要消化医护人员具备相应筛查知识才能正确地选择使用。本研究设计23个消化内镜癌症筛查问题，来评价消化医护人员和新型聊天机器人对消化内镜癌症筛查知识的知晓情况。本研究表明，新必应回答正确率明显高于消化专科医师、ChatGPT、消化专科护士和规培医师，ChatGPT与消化专科医师相当，均优于消化专科护士和规培医师。消化医护人员在此次调查中表现不理想，影响其正确率的独立因素包括人员类别和作答时间。

本研究从以下几个方面获得有意义的发现：(1)既往国内外有较多研究评估消化道癌症筛查参与人群对筛查知识知晓情况[16-21]，本研究是聚焦于消化医护人员对消化内镜癌症筛查知识知晓情况的研究。(2)过去5年中，新消化道癌症筛查指南先后颁布，本研究的问题基于指南制定，发现消化医护人员知晓情况并不理想。(3)新型聊天机器人从2022年11月发布以来，很快被尝试用于心血管疾病预防[22]、美国医师执照考试等[23-25]，表现尚可，本研究评估其在消化内镜癌症筛查相关知识回答方面的应用，发现ChatGPT与消化专科医师表现相当，而新必应表现更佳。ChatGPT训练知识来自2021年以前，不包含2022年以来新颁布的食管癌、胃癌和结直肠癌筛查指南，可能是其表现欠佳的原因之一；而新必应可实时搜索最新互联网信息，并使用更优的GPT4.0构架进行训练，这可能是其表现更佳的原因。尽管目前尚未见新必应在临床医学领域应用的报道，但相关研究必然越来越多。此外，国内诸多类似的聊天机器人(如百度文心一龙等)正在研发或即将发布，期待能在医学问题方面提供更准确、及时的医学知识和服务。

本研究发现消化医护人员对消化内镜癌症筛查问题回答表现欠佳，正确率仅为34.6%~52.2%，且消化专科医师在26.1%(6/23)的相关问题上错误率高达70%以上，具体原因可能包括：(1)本研究筛查问题设置难度过大，39.1%(9/23)是多选题，对消化医护人员的知识水平和看待问题的全面性要求更高。(2)现有筛查指南推广不足，导致消化医护人员对筛查方法和标准不清楚或不统一。(3)消化医护人员知晓情况不佳，提示其对消化内镜癌症筛查缺乏足够重视。(4)未纳入消化专家级医护人员，可能降低了筛查问题回答正确率。(5)少数问题尚缺乏统一标准，可能存在一些争议或不确定性。例如胃黏膜肠化生问题，尽管多个指南认为应该3年进行随访[5, 8]，但有不同观点认为需要根据不同危险分层进行判断[9]。综上所述，有必要从多个层面提高医护人员的总体表现，加强最新筛查指南的宣传和推广，并提供相应培训或辅导，鼓励医护人员主动学习筛查相关知识，加强开展更多筛查相关高质量研究，对部分证据等级低或推荐等级低的问题给出更强的证据和更准确的答案。

在消化医护人员中，消化专科医师的正确率最高，可能原因在于本研究是根据国家颁布的指南和专家共识设置的消化道内镜筛查癌症相关知识问题，具有专业性和学科性特征。相对规培医师和消化专科护士，消化专科医师在所从事的专业和领域有更广泛、更深入的了解和学习，知识储备更夯实。此外，作答时间(≥7.6 min)消化医护人员正确率的独立影响因素，而其在单因素分析时不具有统计学意义，可能在单因素分析时受到其他因素影响，如人员类别中的消化专科护士，72.4%(20名)的消化专科护士回答时间≥7.6 min，其正确率在30%~40%，而其他护士作答时间虽短，正确率却在40%~45%，导致单因素分析时不具有统计学意义。因此，在多因素分析时，需要校正人员类别等因素的影响，才能确定作答时间对提高正确率的独立影响因素。

新型聊天机器人可广泛搜索信息，快速响应问题，但也存在一定的局限性和特殊性，值得进行讨论：(1)新型聊天机器人拒绝回答部分问题。例如ChatGPT和新必应在问题11的100轮测试回答中分别拒绝回答了4次。(2)新必应解释与答案偶有矛盾。新必应有时出现问题解释和答案不一致情

www.zhxhnjzz.com

况，例如在“结肠筛查起始年龄”问题中，新必应多轮回答均指出“不同风险和地区人群的起始年龄可为40岁、45岁和50岁”，但在答案选择时并没有选择40~50岁选项，这提示新必应对问题细节理解仍有不足。(3)对相同问题回答的答案是变化的。值得注意的是，ChatGPT和新必应每一轮回答相同问题的答案可能有差异。我们针对问题11进行100轮测试发现，ChatGPT和新必应的每一轮给出的解释几乎都有描述上的区别，而最终选择的答案也存在较大的异质性(图1)。(4)同时回答多个问题的正确率可能下降：目前ChatGPT对问题的字数并无限制，而新必应要求问题在2 000字以内。本研究23个问题总计约1 500字，ChatGPT和新必应均可一次回答完成，时间均在10 s以内。然而，与逐个回答问题相比，一次回答多个问题时ChatGPT和新必应的正确率似乎有所下降(数据未显示)，这可能与强化学习进行模型训练时的设置不同有关。(5)新聊天机器人回答基于网络大数据，有时易忽视或缺乏临床实践。比如在内镜筛查不良事件中，大数据平台比较关注对人体危害、影响较大的穿孔和出血，而内镜检查术后不适感在临床实践中最常见，术后自行恢复快，对人体危害小，临床实践通常不做特殊处理，对此关注较少，因此ChatGPT和新必应在此问题回答中错误率极高(100.0%)。

此外，随着新必应等智能聊天机器人大规模应用，普通人群获得更准确的消化内镜癌症筛查相关知识更为容易。鉴于当前消化医护人员在筛查问题方面表现欠佳，与普通人群进行交流时可能处于不利地位，因此有必要加强相关知识的学习和获取。有意思的是，本研究中最高得分的一名医师在回答问题过程中也使用新必应作为辅助工具，其正确率超越新必应的平均正确率(82.6% 比 67.0%)，提示人机协同在医疗相关知识问答中有超越单用新型聊天机器人的潜力。

综上，本研究通过问卷和在线聊天的形式评价消化医护人员和新聊天机器人对消化内镜癌症筛查问题的知晓情况，发现消化医护人员对筛查知识的掌握情况尚存在不足。新型聊天机器人特别是新必应回答消化内镜癌症筛查问题具有一定优势，可用于辅助获取消化道癌症筛查的相关知识。然而，其对部分问题的回答仍不满意，仍需专业医师辅助判断。

利益冲突　所有作者声明不存在利益冲突

作者贡献声明　刘亚玲：问卷设计，数据采集，数据统计分析，论文撰写；吕勇、康晨曦、王向平、李静、王陵：数据分析及解释、技术支持；王海英、聂勇战：数据采集、质量监控；吴开春、潘阳林：研究设计，对文章内容进行审阅

参 考 文 献

[1] 北京市科委重大项目《早期胃癌治疗规范研究》专家组. 早期胃癌内镜下规范化切除的专家共识意见(2018,北京)[J]. 中华消化内镜杂志,2019,36(6):381-392.DOI:10.3760/cma.j.issn.1007-5232.2019.06.001.

[2] 国家消化内镜专业质控中心,国家消化系统疾病临床医学研究中心(上海),国家消化道早癌防治中心联盟,等. 中国早期食管癌及癌前病变筛查专家共识意见(2019年,新乡)[J]. 中华消化内镜杂志,2019,36(11):793-801. DOI: 10.3760/cma.j.issn.1007-5232.2019.11.001.

[3] 国家消化系统疾病临床医学研究中心(上海), 国家消化道早癌防治中心联盟, 中华医学会消化内镜学分会, 等. 中国早期结直肠癌筛查流程专家共识意见(2019,上海)[J]. 中华消化内镜杂志, 2019, 36(10):709-719. DOI: 10.3760/cma.j.issn.1007-5232.2019.10.001.

[4] 国家消化系统疾病临床医学研究中心(上海), 中华医学会消化内镜学分会, 中国抗癌协会肿瘤内镜专业委员会, 等. 中国结直肠癌癌前病变和癌前状态处理策略专家共识[J]. 中华消化内镜杂志, 2022, 39(1):1-18. DOI: 10.3760/cma.j.cn321463-20211111-00661.

[5] 赫捷,陈万青,李兆申,等. 中国胃癌筛查与早诊早治指南(2022,北京)[J]. 中华肿瘤杂志,2022,44(7):634-666. DOI:10.3760/cma.j.cn112152-20220617-00430.

[6] 中国抗癌协会大肠癌专业委员会中国结直肠肿瘤早诊筛查策略制订专家组. 中国结直肠肿瘤早诊筛查策略专家共识[J]. 中华胃肠外科杂志,2018,21(10):1081-1086. DOI:10.3760/cma.j.issn.1671-0274.2018.10.001.

[7] 中国抗癌协会胃癌专业委员会. CACA胃癌整合诊治指南(精简版)[J]. 中国肿瘤临床,2022,49(14):703-710. DOI:10.12354/j.issn.1000-8179.2022.20220712.

[8] 中华人民共和国国家卫生健康委员会医政医管局. 胃癌诊疗指南(2022年版)[J]. 中华消化外科杂志,2022,21(9):1137-1164.DOI:10.3760/cma.j.cn115610-20220726-00432.

[9] 中华医学会肿瘤学分会,中华医学会杂志社. 中华医学会胃癌临床诊疗指南(2021版)[J]. 中华医学杂志,2022,102(16):1169-1189.DOI:10.3760/cma.j.cn112137-20220127-00197.

[10] 中华医学会肿瘤学分会早诊早治学组. 中国结直肠癌早诊早治专家共识[J]. 中华医学杂志,2020,100(22):1691-1698. DOI:10.3760/cma.j.cn112137-20190924-02103.

[11] 中华医学会肿瘤学分会早诊早治学组. 中国食管癌早诊早治专家共识[J]. 中华肿瘤杂志,2022,44(10):1066-1075.DOI:10.3760/cma.j.cn112152-20220220-00114

[12] Zhang L, Mosquera I, Lucas E, et al. CanScreen5, a global repository for breast, cervical and colorectal cancer screening programs[J]. Nat Med, 2023, 29(5):1135-1145. DOI: 10.1038/s41591-023-02315-6.

[13] Chen H, Li N, Ren J, et al. Participation and yield of a population-based colorectal cancer screening programme in China[J]. Gut, 2019, 68(8): 1450-1457. DOI: 10.1136/gutjnl-2018-317124.

[14] Wang H, Wu W, Dou Z, et al. Performance and exploration of ChatGPT in medical examination, records and education in Chinese: pave the way for medical AI[J]. Int J Med Inform, 2023, 177:105173. DOI: 10.1016/j.ijmedinf.2023.105173.

www.zhxhnjzz.com

[15] Takagi S, Watari T, Erabi A, et al. Performance of GPT-3.5 and GPT-4 on the Japanese medical licensing examination: comparison study[J]. JMIR Med Educ, 2023, 9:e48002. DOI: 10.2196/48002.
[16] Tran MT, Jeong MB, Nguyen VV, et al. Colorectal cancer beliefs, knowledge, and screening among Filipino, Hmong, and Korean Americans[J]. Cancer, 2018, 124(Suppl 7): 1552-1559. DOI:10.1002/cncr.31216.
[17] Duron V, Bii J, Mutai R, et al.Esophageal cancer awareness in Bomet district, Kenya[J]. Afr Health Sci, 2013, 13(1): 122-128. DOI:10.4314/ahs.v13i1.17.
[18] Liu Q, Zeng X, Wang W, et al. Awareness of risk factors and warning symptoms and attitude towards gastric cancer screening among the general public in China: a cross-sectional study[J]. BMJ Open, 2019, 9(7): e029638. DOI: 10.1136/bmjopen-2019-029638.
[19] 冯祥,朱进华,宋统球,等.江苏省扬中市上消化道癌高危人群筛查满意度调查[J].中华肿瘤防治杂志,2022,29(8):549-553,570. DOI: 10.16073/j.cnki.cjcpt.2022.08.04.
[20] 滕菲, 李贺, 曹毛毛, 等.基于中国滕菲,李贺,曹毛毛,等.基于中国部分农村上消化道癌筛查地区的癌症防治核心知识知晓情况分析[J].肿瘤预防与治疗,2020,33(6):493-500. DOI: 10.3969/j.issn.1674-0904.2020.06.007.
[21] 周荣耀,沈兴蓉,冯瑞,等.合肥市居民上消化道癌筛查意愿的心理认知因素分析[J].南京医科大学学报(社会科学版),2018,18(1):25-29. DOI:10.7655/NYDXBSS20180107.
[22] Sarraju A, Bruemmer D, Van Iterson E, et al. Appropriateness of cardiovascular disease prevention recommendations obtained from a popular online chat-based artificial intelligence model[J]. JAMA, 2023, 329(10): 842-844. DOI: 10.1001/jama.2023.1044.
[23] Cascella M, Montomoli J, Bellini V, et al. Evaluating the feasibility of ChatGPT in healthcare: an analysis of multiple clinical and research scenarios[J]. J Med Syst, 2023, 47(1): 33. DOI:10.1007/s10916-023-01925-4
[24] Huh S. Are ChatGPT′s knowledge and interpretation ability comparable to those of medical students in Korea for taking a parasitology examination?: a descriptive study[J]. J Educ Eval Health Prof, 2023,20:1. DOI:10.3352/jeehp.2023.20.1
[25] Gilson A, Safranek CW, Huang T, et al. How does ChatGPT perform on the united states medical licensing examination? The implications of large language models for medical education and knowledge assessment[J]. JMIR Med Edu, 2023, 9:e45312. DOI:10.2196/45312.

·读者·作者·编者·

发表学术论文“五不准”

1.不准由“第三方”代写论文。科技工作者应自己完成论文撰写,坚决抵制“第三方”提供论文代写服务。

2.不准由“第三方”代投论文。科技工作者应学习、掌握学术期刊投稿程序,亲自完成提交论文、回应评审意见的全过程,坚决抵制“第三方”提供论文代投服务。

3.不准由“第三方”对论文内容进行修改。论文作者委托“第三方”进行论文语言润色,应基于作者完成的论文原稿,且仅限于对语言表达方式的完善,坚决抵制以语言润色的名义修改论文的实质内容。

4.不准提供虚假同行评审人信息。科技工作者在学术期刊发表论文如需推荐同行评审人,应确保所提供的评审人姓名、联系方式等信息真实可靠,坚决抵制同行评审环节的任何弄虚作假行为。

5.不准违反论文署名规范。所有论文署名作者应事先审阅并同意署名发表论文,并对论文内容负有知情同意的责任;论文起草人必须事先征求署名作者对论文全文的意见并征得其署名同意。论文署名的每一位作者都必须对论文有实质性学术贡献,坚决抵制无实质性学术贡献者在论文上署名。

本“五不准”中所述“第三方”指除作者和期刊以外的任何机构和个人;“论文代写”指论文署名作者未亲自完成论文撰写而由他人代理的行为;“论文代投”指论文署名作者未亲自完成提交论文、回应评审意见等全过程而由他人代理的行为。

www.zhxhnjzz.com

麦康医疗

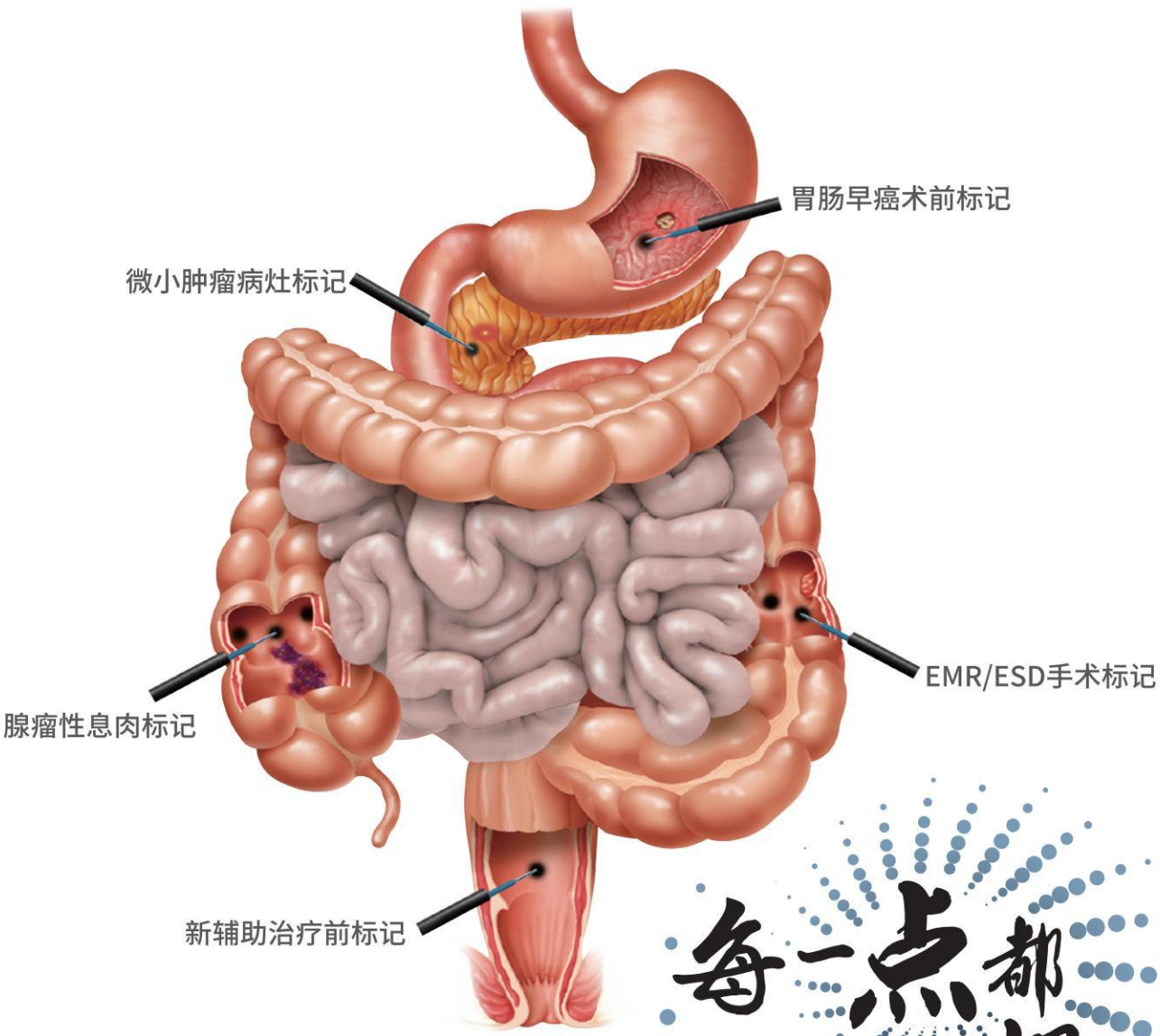
广告

Spot 内镜定位标记液

ENDOSCOPIC MARKER



适用于内镜下对胃肠道病灶进行标记或染色



每一点都恰到好处

快速确定病灶位置,避免误切

减少病灶探查时间,节省手术时间

辅助判断肿瘤边界,更小的有效切除范围

北京麦康医疗器械有限公司
BEIJING MEDKON MEDICAL DEVICES CO., LTD

产品名称：内镜定位标记液
注册证编号：国械注进 20223060062
生产厂商：GI Supply Inc.
代 理 人：北京麦康医疗器械有限公司

电　　话：010-85894489 转 837
地　　址：北京市朝阳区建国路 88 号现代城 B 座 806 室
禁　　忌：禁忌内容或者注意事项详见说明书
京械广审（文）第 270210-12521 号

健可诺®
国药准字H20213838

磷酸钠盐散
Sodium Phosphates Powder

广告

清肠品质卓越 舒适又方便

独特包装
控制用药风险



适 应 症 用于患者结肠X-光线及肠道内窥镜检查前或手术前清理肠道。

用法用量 本品用于肠道准备时服药一般分两次，每次服药1袋。
第一次服药时间在手术或检查前一天晚上7点，用法为用800ml以上温凉水溶解后服用。
第二次服药时间在手术或检查当天早上7点(或在操作或检查前至少3个小时)，或遵医嘱，用法同第一次。
为获得良好肠道准备效果，建议患者在可承受范围内多饮用水。

不良反应 常见的不良反应为腹胀、恶心、腹痛、呕吐，还可能会出现用药期间和用药后的短暂的电解质紊乱、乏力、眩晕、过敏反应、肝功能检查ALT，AST升高、肛门刺激症状。其他详见说明书。

禁 忌 1.本品禁用于先天性巨结肠、肠梗阻、腹水、充血性心脏病或肾功能衰竭患者。2.使用本品禁止联合使用其他缓泻药物。3.对本品中任何成份过敏者禁用。

健能药业 JEWELLAND

川药广审(文)第251011-01618号 生产企业:四川健能制药有限公司 本广告仅供医学药学专业人士阅读