



人工智能随访和人工随访的一致性研究： 以高血压共病糖尿病患者为例*

——吴静¹ 王思源² 陈秋艳¹ 王玉恒² 杨沁平² 张晟² 徐璐璐² 吴萃¹ 程旻娜²

【摘要】 **目的** 评价人工智能(AI)随访和人工随访在高血压共病糖尿病患者中的一致性。**方法** 于2021年7月17日—12月10日,对981名在管高血压共病糖尿病患者分别进行人工随访和AI随访。比较两种随访方式的异常报告率,并采用Kappa方法分析随访结果一致性。**结果** AI随访接通率为90.09%,信息采集率为70.72%,各年龄组间接通率差异具有统计学意义($P=0.047$),双休日的接通率和采集率均显著高于工作日($P<0.001$),拨打时段中9:00~12:00的采集率高于其他时段($P=0.010$)。两种随访方式在“烦躁”“面色苍白或潮红”和“不规律活动”条目中的异常报告率差异具有统计学意义($P<0.05$)。AI随访和人工随访在症状类条目中的一致性水平在较差和高度之间波动,在行为类条目中的一致性水平较高。**结论** AI随访可用于慢性病健康管理。未来需绘制患者数字画像,定制个性化随访计划;优化AI随访内容,采用AI随访与人工随访相结合方式。

【关键词】 慢性病;高血压;糖尿病;共病;人工智能随访;电话随访;随访一致性

中图分类号:R197.323

文献标识码:A

Consistency Study of Artificial Intelligence Follow-up and Manual Follow-up: a Case Study of Hypertensive Patients with Diabetes Mellitus/WU Jing, WANG Siyuan, CHEN Qiuyan, et al. // Chinese Health Quality Management, 2025, 32(3): 92-96

Abstract **Objective** To evaluate the consistency of artificial intelligence (AI) follow-up and manual follow-up in patients with hypertension and diabetes mellitus. **Methods** From July 17 to December 10, 2021, 981 patients with hypertension and diabetes mellitus were followed-up manually and by AI respectively. The abnormal report rate of the two follow-up methods were compared, and the consistency of the follow-up results was analyzed by Kappa method.

Results The follow-up connection rate of AI was 90.09%, and the collection rate was 70.72%. There difference in the connection rate among different age groups was statistically significant ($P=0.047$). The connection rate and collection rate on weekends were significantly higher than those on weekdays ($P<0.001$). The collection rate from 9:00 to 12:00 was higher than that in other periods ($P=0.010$). There were statistically significant differences in the abnormal report rate of the two follow-up methods in the items of "irritability", "pale or flush" and "irregular activity" ($P<0.05$). The consistency level of AI follow-up and manual follow-up in symptom items fluctuated between poor and high, with the level in behavior items being higher. **Conclusion** AI follow-up can be used for chronic disease health management. In the future, it is necessary to draw digital portraits of patients and customize personalized follow-up plans. AI follow-up content needs to be optimized, with a combined method of AI follow-up and manual follow-up.

Key words Chronic Disease; Hypertension; Diabetes; Comorbidity; Artificial Intelligence Follow-up; Telephone Follow-up; Consistency of Follow-up

First-author's address Shanghai Baoshan District Center for Disease Control and Prevention, Shanghai, 201901, China

作为社区居民中常见的慢性病 称“高糖共病”)不仅影响了患者的 负担^[1-2]。《健康中国行动2023年 共患病,高血压共病糖尿病(以下简 生活质量,而且加重了患者的疾病 工作要点》^[3]指出,要加强高血压、

DOI:10.13912/j.cnki.chqm.2025.32.3.19

* 基金项目:上海市公共卫生体系建设三年行动计划(2023—2025年)(编号:GWVI-8)

1 上海市宝山区疾病预防控制中心 上海 201901 2 上海市疾病预防控制中心慢性病与伤害防治所 上海 200336

糖尿病等慢性病患者的健康管理,推进医防融合,提升服务质量。随访作为慢性病患者健康管理的重要一环,不仅有助于收集患者健康数据、监测病情变化、评估治疗效果,而且能有效降低并发症发生风险^[4]。传统随访主要由医护人员通过电话或面对面等方式进行,虽然能够为患者提供情感支持和人文关怀,但由于基层医疗卫生机构中负责慢性病管理的医务人员数量较少,加之慢性病管理工作量大,导致随访效率低下,且无法保障随访过程的规范性。近年来,人工智能(Artificial Intelligence, AI)在医疗领域的应用日益广泛,如体检预约、健康教育、出院患者随访等^[5],提高了医务人员的工作效率。目前,AI随访相关研究集中在系统构建与应用方面^[6-7],关于两种随访方式的一致性研究鲜有报道。基于此,本研究分析了AI随访和人工随访在高糖共病患者中的一致性,旨在优化AI随访模式,提高患者随访依从性和满意度,从而推动AI技术在慢性病患者健康管理中的应用。

1 对象与方法

1.1 研究对象

2021年7月17日—12月10日,选择宝山区罗店、顾村和大场三个社区卫生服务中心在管的高糖共病患者981例。纳入标准:(1)年龄 ≥ 35 岁;(2)已确诊为高血压和糖尿病共病患者;(3)有一定沟通能力,且同意接受AI随访和医生电话随访。排除标准:(1)合并严重的心肝肾等脏器疾病;(2)精神神经及行为异常者;(3)基本信息采集错误。

1.2 研究方法

1.2.1 随访内容

采用《上海市社区健康管理工

作规范——慢性病综合防治(2017年版)》中的“管理随访信息表”收集随访资料,共有23个与高血压和糖尿病相关的问题,分为症状类和行为类。症状类问题包括头晕头痛、恶心呕吐、乏力、烦躁、面色苍白或潮红、四肢发麻或下肢水肿或肢端溃疡、视力模糊或耳鸣眼花、鼻出血、皮肤瘙痒、胸闷心悸或呼吸困难、低血糖、多食、多饮、多尿、感染疾病、有并发症、药物不良反应等;行为类问题包括未接受健康教育、摄盐过度、吸烟、饮酒、不遵医嘱服药、不规律活动等。

1.2.2 随访方法

所有研究对象均需接受医生电话随访和AI随访,两次随访内容相同,间隔时间为3d~5d。

AI随访采用阿里云提供的语音识别技术和上海市疾病预防控制中心开发的语音随访平台。工作流程主要分为四个部分:(1)输入患者信息。首先导入每个随访对象的基本信息列表,包括身份证号码、电话号码、主治医师等。(2)拨打电话。平台自动通过中继线路进行拨号,并根据编制好的随访内容使用模拟人声进行随访。在9:00~12:00、12:00~17:00、17:00~18:00任一时间段内进行拨打,若该时段患者未接听,则在后一时间段再次拨打,若三次未接听,则该随访周期内不再拨打。

(3)提问和收集信息。收到患者回答后,利用语音识别技术和自然语言处理技术提取语音数据要素,并通过语音合成、语音识别技术、语音模式识别等生成标准化指标。(4)保存随访内容及反馈指标。将生成的标准化指标传输至慢性病管理信息系统,同时将随访日期、呼叫和应答情况、随访时间、完整录音、片段录音、翻译文本、随访结果等进行结构化存储。AI随访流程见图1。

人工随访由经过专业培训的社区医生完成,随访结果由医生手工录入电脑,同时通话被录音,便于比对随访结果,进一步完善随访话术和流程。

1.3 评价指标

AI呼叫接通率=呼叫成功人数/计划拨打人数 $\times 100\%$ 。信息采集率=至少回答一题人数/接通人数 $\times 100\%$ 。异常报告率=某条目报告存在异常症状或行为人数/回答该条目总人数 $\times 100\%$ 。一致性分析采用Kappa值(κ)计算^[8],具体分类和水平如下:0.00~0.20(较差),0.21~0.40(一般),0.41~0.60(中等),0.61~0.80(高度),0.81~1.00(极好)。

1.4 统计分析方法

使用SPSS26.0软件进行统计

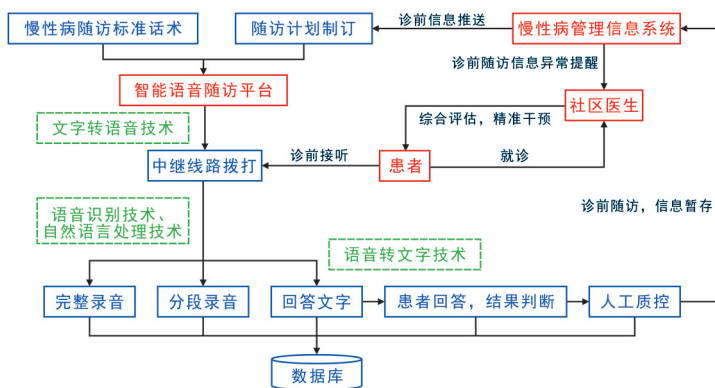


图1 人工智能随访流程

分析。呈偏态分布的计量资料用中位数和四分位数间距 [$M(P_{25}, P_{75})$] 表示, 组间比较采用秩和检验。计数资料采用频数和率表示, 组间比较采用卡方检验。两种随访方式间异常报告率的比较采用配对卡方检验, 率的趋势变化采用趋势卡方检验。检验水准 $\alpha=0.05$ 。

2 结果

2.1 研究对象基本情况

在招募的 981 例高糖共病患者中, 去除基本信息采集错误者, 共对 959 例患者进行人工随访和 AI 随访。其中: 男性 497 例(51.82%), 女性 462 例(48.18%); 中位年龄 69(62,75)岁, 以 65 岁~<75 岁患者居多, 共 389 例(40.56%)。研究对象基本情况见表 1。

2.2 AI 随访接通和信息采集情况

在 AI 随访的 959 例高糖共病患者中, AI 呼叫成功 864 例, 接通率为 90.09%。各年龄组间接通率差异具有统计学意义($\chi^2=3.946, P=0.047$), 在 ≥ 45 岁患者中, 接通率随年龄增加而上升。双休日接通率明显高于工作日, 差异具有统计学意义($\chi^2=10.344, P<0.001$)。见表 1。

去除接通未采集信息者, AI 随访总体采集率为 70.72% (611/864)。双休日采集率显著高于工作日($\chi^2=14.501, P<0.001$), 拨打时段中 9:00~12:00 的采集率高于其他时段($\chi^2=6.554, P=0.010$)。见表 1。

在采集到随访信息的 611 例患者中, 平均回答题数为 20(16,22)个, 且受过中高等教育的患者回答题数显著多于小学及以下学历患者($H=33.335, P<0.001$)。见表 1。

2.3 AI 随访和人工随访的异常报告率情况

AI 随访中, 各条目的采集率随条目序次增加而降低, 差异具有统计学意义($\chi^2_{趋势}=33.316, P<0.001$)。在人工随访和 AI 随访均采集到的条目中, “未接受健康教育” 条目异常报告率均最高, 人工随访为 83.23%、AI 随访为 82.80%。人工随访中, “不遵医嘱服药” 条目异常报告率最低, 为 1.90%; AI 随访中, “鼻出血” 条目异常报告率最低, 为 1.16%。此外, 人工随访和

AI 随访在“烦躁”“面色苍白或潮红”“不规律活动”三个条目中的异常报告率差异具有统计学意义($P<0.05$), 见表 2。

2.4 AI 随访和人工随访的一致性情况

在症状方面, AI 随访和人工随访的一致性水平在较差和高度之间波动, “低血糖” 条目的一致性最高($\kappa=0.615$), 除“药物不良反应” 条目外, 其余条目一致性差异均有统计学意义($P<0.001$)。在行为方面, AI 随

表 1 高糖共病患者 AI 随访接通和信息采集情况

项目	接通情况($n=864$)			信息采集情况($n=611$)			信息采集题数($n=23$)		
	接通人数 (占比)/ 例(%)	χ^2	P	采集人数 (占比)/ 例(%)	χ^2	P	题数/个	H	P
性别									
男($n=497$)	445 (89.54)	0.358	0.550	305 (68.54)	2.103	0.147	20 (17, 22)	0.044	0.834
女($n=462$)	419 (90.69)			306 (73.03)			20 (16, 22)		
年龄/岁									
35~<45 ($n=19$)	19 (100.00)	3.946	0.047	13 (68.42)	0.703	0.402	22 (16, 22)	9.160	0.057
45~<55 ($n=73$)	62 (84.93)			43 (69.35)			21 (13, 22)		
55~<65 ($n=229$)	198 (86.46)			137 (69.19)			20 (16, 22)		
65~<75 ($n=389$)	353 (90.75)			249 (70.54)			20 (18, 22)		
≥ 75 ($n=249$)	232 (93.17)			169 (72.84)			19 (15, 21)		
学历									
小学及以下 ($n=167$)	150 (89.82)	0.663	0.416	109 (72.67)	0.688	0.407	19 (14, 21)	33.335	<0.001
初中/中专 ($n=423$)	383 (90.54)			277 (72.32)			20 (18, 22)		
高中/高职 ($n=167$)	147 (88.02)			103 (70.07)			21 (18, 22)		
大专及以上 ($n=159$)	143 (89.94)			101 (70.63)			21 (17, 22)		
不详 ($n=43$)	41 (95.35)			21 (51.22)			12 (3, 19)		
拨打时间									
工作日 ($n=624$)	548 (87.82)	10.344	<0.001	363 (66.24)	14.501	<0.001	20 (16, 22)	3.707	0.054
双休日 ($n=335$)	316 (94.33)			248 (78.48)			21 (17, 22)		
拨打时段									
9:00~12:00 ($n=437$)	398 (91.08)	2.699	0.100	299 (75.13)	6.554	0.010	20 (16, 21)	3.428	0.180
12:00~17:00 ($n=272$)	249 (91.54)			169 (67.87)			21 (17, 22)		
17:00~18:00 ($n=250$)	217 (86.80)			143 (65.90)			20 (15, 22)		

注: 拨打时段在 12:00 或 17:00 时归属于前组。

访和人工随访的一致性水平较高,尤其是“吸烟”(κ=0.891)和“饮酒”(κ=0.699),除“不遵医嘱服药”条目外,其余条目一致性差异均有统计学意义($P<0.001$)。见表 3。

3 讨论

3.1 绘制患者数字画像,定制化随访计划

在接通率方面,总体来看,本研究结果显示,AI 随访接通率为 90.09%,高于谷孝云等^[9]在基本公共卫生服务领域试点应用的两次重拨接通率 67.30%,以及夏远亲等^[10]对体检人群检后随访的接通率 75%。究其原因,本研究在医生电话随访的同时会再次告知患者近一周内将会进行 AI 随访。具体来看,AI 随访接通率在中老年人群中随年龄增加而上升,可能与此类人群可自由支配时间较充足有关。

在采集率方面,本研究结果显示,AI 随访各条目的采集率随条目序次增加而降低($P<0.001$),且受过中高等教育的患者答题数显著多于小学及以下学历患者。这在一定程度上表明,文化程度越高,患者对 AI 随访的接受度越高。对此,建议加强宣传教育,以提高患者 AI 随访依从性。

本研究结果还发现,无论是接通率还是采集率,均为双休日高于工作日。但有研究^[7]显示,在上海市黄浦区打浦桥社区,AI 随访工作日的采集率高于双休日,可能与患者群体不同有关。这提示,关于随访时间的设置不能一概而论,应根据实际制订具体方案,以确保应答率,提高随访效率。此外,本研究对拨打时段进行了细致划分,结果显示各时段 AI 随访的接通率均>85%,且上午 9:00~12:00 时段的

表 2 高糖共病患者 AI 随访和人工随访各条目异常报告率情况

序次	条目	共同采集数/例	异常报告人数(占比)/例(%)			
			人工随访	AI 随访	χ^2	P
1	头晕头痛	567	93 (16.40)	76 (13.40)	2.943	0.086
2	恶心呕吐	564	18 (3.19)	17 (3.01)	0.000	1.000
3	乏力	546	130 (23.81)	149 (27.29)	2.473	0.116
4	烦躁	541	86 (15.90)	143 (26.43)	24.693	<0.001
5	面色苍白或潮红	538	31 (5.76)	48 (8.92)	5.698	0.017
6	四肢发麻或下肢水肿或肢端溃疡	542	147 (27.12)	136 (25.09)	0.917	0.338
7	视力模糊或耳鸣眼花	139	55 (39.57)	60 (43.17)	0.485	0.486
8	鼻出血	516	12 (2.33)	6 (1.16)	1.786	0.180
9	皮肤瘙痒	532	191 (35.90)	209 (39.29)	2.258	0.133
10	胸闷心悸或呼吸困难	522	83 (15.90)	78 (14.94)	0.232	0.630
11	低血糖	488	61 (12.50)	51 (10.45)	1.841	0.175
12	多尿	484	73 (15.08)	76 (15.70)	0.049	0.824
13	感染疾病	514	21 (4.09)	27 (5.25)	0.962	0.327
14	有并发症	488	25 (5.12)	15 (3.07)	2.893	0.089
15	未接受健康教育	471	392 (83.23)	390 (82.80)	0.012	0.913
16	摄盐过度	421	34 (8.08)	44 (10.45)	2.025	0.155
17	多食	453	69 (15.23)	78 (17.22)	0.928	0.336
18	多饮	436	113 (25.92)	121 (27.75)	0.408	0.523
19	吸烟	470	84 (17.87)	79 (16.81)	1.067	0.302
20	饮酒	458	77 (16.81)	72 (15.72)	0.432	0.511
21	不遵医嘱服药	263	5 (1.90)	10 (3.80)	1.067	0.302
22	药物不良反应	397	26 (6.55)	15 (3.78)	2.703	0.100
23	不规律活动	417	135 (32.37)	111 (26.62)	6.612	0.010

表 3 高糖共病患者 AI 随访和人工随访的一致性分析(n=415)

类别	条目	Kappa 值(κ)	等级	P
症状类	低血糖	0.615	高度	<0.001
	视力模糊或耳鸣眼花	0.534	中等	<0.001
	皮肤瘙痒	0.513	中等	<0.001
	四肢发麻或下肢水肿或肢端溃疡	0.485	中等	<0.001
	胸闷心悸或呼吸困难	0.478	中等	<0.001
	乏力	0.448	中等	<0.001
	多食	0.431	中等	<0.001
	感染疾病	0.416	中等	<0.001
	面色苍白或潮红	0.394	一般	<0.001
	多尿	0.386	一般	<0.001
	头晕头痛	0.364	一般	<0.001
	恶心呕吐	0.313	一般	<0.001
	烦躁	0.303	一般	<0.001
行为类	多饮	0.302	一般	<0.001
	鼻出血	0.298	一般	<0.001
	有并发症	0.197	较差	<0.001
	药物不良反应	0.090	较差	0.062
	吸烟	0.891	极好	<0.001
	饮酒	0.699	高度	<0.001
	不规律活动	0.541	中等	<0.001
	摄盐过度	0.410	中等	<0.001
	未接受健康教育	0.390	一般	<0.001
	不遵医嘱服药	-0.027	—	0.648

注:对两种随访方式答题数均≥19 题者进行分析。

信息采集率显著高于其他时段。

综上,建议智能语音随访平台根据患者的年龄、学历以及拨打时间、时段和次数等信息,深度挖掘患者的特征和偏好,绘制患者数字画像,将患者按不同属性分类,并针对不同类型患者定制个性化随访计划,从而确保随访工作的针对性和有效性。

3.2 优化 AI 随访内容,采用 AI 随访与人工随访相结合方式

本研究发现,症状类条目的随访一致性水平波动较大,与Mohammed A等^[11]研究结果一致,这可能与患者对条目的理解不够深入,在两次随访中的回答存在偏差有关。具体来看,在烦躁、面色苍白或潮红等症状方面,AI随访的异常报告率高于人工随访,两者的一致性处于一般水平;而低血糖、皮肤瘙痒、胸闷心悸等症状类条目的随访一致性处于中等及以上水平。由此可见,症状具有主观性。一项在孟加拉国和坦桑尼亚两个国家开展的研究^[12]表明,饮酒和吸烟在两种随访方式间的一致性较高,这与本研究结果一致,且两种随访方式在吸烟和饮酒方面的异常报告率差异无统计学意义,可能与行为具有一定的稳定性有关。对此,建议在每年对高糖共病患者的四次随访中,减少对吸烟和饮酒的随访次数,如半年或一年随访一次。

为了进一步分析导致人工随访和AI随访结果不一致的原因,本研究随机听取了部分AI随访录音,总结了以下四种情况:(1)回答不一致(占71.81%)。即两次随访回答存在差异。(2)判别不一致(占7.72%)。患者回答“偶尔有”或“有时有”时,AI识别为有,但医生可能识别为没有。(3)AI识别错误(占3.02%)。如患者回答“有一点”,AI

识别为没有。(4)医生识别错误(占12.75%)。如患者回答“正常”,医生识别为有,不排除无意识填写错误。针对第一种情况,可能与多数高糖共病患者年龄较大,记忆力和理解力下降有关。针对后三种情况,原因可能有四点:第一,AI只能根据既定话术进行提问,无法与患者正常交流,而人工随访中医生会深入追问;第二,AI不能很好地判断模棱两可的回答;第三,两次随访接听电话者可能不同;第四,随访话术中每道题目的回答时间有限。综上,建议:一是通过与用户不断交互,提高AI处理能力;二是对AI随访内容进行分类,并根据不同类型的问题设定随访周期,如减少对生活行为类问题的随访频率;三是根据每道题目的长短和复杂程度,针对不同人群调整预留的答题时间;四是对随访话术模板外的问题,AI尚不能有效解决时可通过人工进行随访。

注:吴静、王思源为共同第一作者。

参考文献

- [1] 孙文岩,王思源,吴菲,等.上海市35岁及以上居民高血压、糖尿病共病流行现状分析[J].上海预防医学,2021,33(7):559-565.
- [2] 蔡仕良,鲜林峰,李子帅,等.上海郊区居民高血压和糖尿病共患的相关因素分析[J].上海预防医学,2022,34(10):1033-1037.
- [3] 健康中国行动推进委员会办公室.关于印发健康中国行动2023年工作要点的通知:国健推委办发〔2023〕1号[EB/OL].(2023-03-06)[2024-08-01].<http://www.nhc.gov.cn/guihuaxxs/s7788/202303/1a6c05141f9547768ba1f7792c0313d0.shtml>.
- [4] 韩帅,朱雪娇,舒洁,等.慢性病患者云随访服务体验量表的研制及其信效度检验[J].中国全科医学,2024,27(28):

3552-3559.

[5] 程美,王力华.医疗智能语音技术与应用综述[J].中国数字医学,2021,16(8):1-7.

[6] 黄一敏,王森,魏明月,等.人工智能语音随访系统在儿童日间手术中心的应用[J].中国卫生质量管理,2021,28(3):9-11.

[7] 王思源,周峰,高俊岭,等.人工智能电话随访在高血压随访管理中的应用[J].中国慢性病预防与控制,2021,29(11):817-820.

[8] MCHUGH ML. Interrater reliability: the kappa statistic[J]. Biochem Med (Zagreb), 2012,22(3):276-282.

[9] 谷孝云,杨效军,谷张鹏.智能语音外呼系统在公共卫生服务中的应用实践[J].中国卫生信息管理杂志,2023,20(2):285-289.

[10] 夏远亲,王建刚,王桂莲,等.健康管理智能语音检后随访平台的构建[J].医疗装备,2023,36(17):26-28,40.

[11] MOHAMMED A, ACHEAMPONG PR, OTUPIRI E, et al. Symptom monitoring of childhood illnesses and referrals: a pilot study on the feasibility of a mobile phone-based system as a disease surveillance tool in a rural health district of Ghana[J]. Health Informatics J, 2020,26(2):1465-1476.

[12] PARIYO GW, GREENLEAF AR, GIBSON DG, et al. Does mobile phone survey method matter? Reliability of computer-assisted telephone interviews and interactive voice response non-communicable diseases risk factor surveys in low and middle income countries [J]. PLoS One, 2019, 14(4): e0214450.

通信作者:

吴萃:上海市宝山区疾病预防控制中心慢一科科长

E-mail:25850569@qq.com

程旻娜:上海市疾病预防控制中心慢性病与伤害防治所健康管理科科长

E-mail:chengminna@sdc.sh.cn

收稿日期:2024-09-06

修回日期:2024-10-08

责任编辑:任红霞