

1 **A Novel Artificial Intelligence Platform to Automate Clinical Consultation Notes and**  
2 **Enhance Diagnostic Efficiency in the Outpatient Clinic: Proposal of a Protocol for a Multi-**  
3 **Center, Multi-Disciplinary, Prospective Randomized Controlled Trial**

4  
5  
6  
7 Karanvir Gill\*<sup>1</sup>, Giovanni Cacciamani<sup>1</sup>, Jamal Nabhani\*<sup>1</sup>, Joshua Corb\*<sup>2</sup>, Tom Buchanan<sup>1</sup>, Daniel  
8 Park<sup>1</sup>, Virinder Bhardwaj<sup>3</sup>, Onkarjit Marwah<sup>4</sup>, Moses Kim<sup>5</sup>, Deepak Kapoor<sup>6</sup>, Alexander Kutikov<sup>7</sup>,  
9 Robert Uzzo<sup>7</sup>, Inderbir Gill\*<sup>#1</sup>

10  
11  
12  
13  
14 1- Catherine & Joseph Aresty Department of Urology, USC Institute of Urology, University of  
15 Southern California, Los Angeles, CA; 2 - Basis Worldwide; 3- Department of Urology, Sierra  
16 View Medical Center, Porterville, CA; 4- Heartbeat Cardiovascular Medical Group, Glendale,  
17 CA, 5- Orange Coast Urology Associates, Newport Beach, CA; 6- Solaris Urology Associates,  
18 New York, NY; 7- Department of Urology, Fox Chase Cancer Center, Philadelphia, PA  
19  
20  
21  
22  
23  
24  
25  
26

27 Words: Abstract 434

28 Text: 4699

29 Tables: 1

30 References: 24

31  
32  
33  
34 \*Conflict of Interest: OneLine Health  
35  
36  
37

38 #Correspondence:

39 Inderbir S. Gill, MD

40 Catherine & Joseph Aresty Department of Urology

41 Keck School of Medicine, University of Southern California 1441 Eastlake Ave, Suite 7416

42 Los Angeles, CA 90089

43 Email: [igill@med.usc.edu](mailto:igill@med.usc.edu); [gillindy@gmail.com](mailto:gillindy@gmail.com)

44 Cell: 216-312-6700  
45  
46  
47  
48  
49  
50  
51

52 **ABSTRACT**

53 Presented herein is a proposal for a protocol for a multi-center, multi-disciplinary randomized  
54 controlled trial (RCT) to evaluate a novel artificial intelligence (AI)-based technology that  
55 automates the construction of the clinical consultation note (CCN) and enhances diagnostic  
56 assessments in the outpatient clinic setting. This innovative tech-platform automatically  
57 generates the CCN and presents it to the provider in advance of the patient consultation,  
58 without any work done by the provider. The constructed CCN is presented either in the native  
59 electronic health record (EHR) or in a secure web-based application, in a HIPAA-compliant  
60 manner. The proposed prospective trial will compare this novel AI/ML technology  
61 (NAMT) versus the current standard-of-care (SOC) in the outpatient setting. Outpatient clinic-  
62 days will be randomized to either “SOC clinic-day” or the “NAMT clinic-day” based on whether  
63 the SOC or the NAMT was used to construct the CCN for all patients seen on that particular  
64 clinic-day. Randomized cross-over of each provider between “SOC clinic-day” and “NAMT clinic-  
65 day” will result in each provider serving as her/his own internal control. Objective data will be  
66 used to compare study endpoints between the SOC and the NAMT. Co-primary endpoints  
67 include a) CCN diagnostic accuracy/quality (based on standardized QNOTE metrics); and b)  
68 Work-outside-work (WOW) time required by providers to complete clinic-related documentation  
69 tasks outside clinic hours (based on EHR meta-data). Secondary endpoints include a) Provider  
70 productivity (based on provider “walk-in, walk-out” time from the consultation room); b) Provider  
71 satisfaction (based on the standardized AHRQ EHR End User Survey); and c) Patient  
72 satisfaction (based on the standardized Press Ganey/CG-CAHPS survey). To assess  
73 generalizability across the health-care spectrum, the study will be conducted in four different  
74 types of health-care settings (large academic medical center; non-academic hospital; rural  
75 hospital; community private practice); in four different disciplines (cardiology; infectious disease;  
76 urology; emergency medicine); using four different EHR systems (Cerner; Epic; AllScripts;  
77 MediTech/UroChart). We estimate an aggregate RCT sample size of 150 clinic-days (involving

78 3,000 total patients; 15-30 providers). This will randomize 75 clinic-days (1,500 patients) to the  
79 control SOC arm, and 75 clinic-days (1,500 patients) to the intervention NAMT arm. We will use  
80 a two-sided Z-test of difference between proportions with 90% power and two-sided 5%  
81 significance level. This RCT is the first to evaluate the efficiency and diagnostic accuracy of pre-  
82 constructing CCNs in an automated manner using AI/ML technology, deployed at a large-scale,  
83 multi-institutional, multi-disciplinary, multi-EHR level. Results from this study will provide  
84 definitive level 1 evidence about the desirability and generalizability of AI-generated  
85 automatically constructed CCNs, assessing its potential benefits for providers, patients, and  
86 healthcare systems.

87

88

89

90

91

92

93

94

95

96

97

98

99

100

101

102

103

## 104 **INTRODUCTION**

105 Electronic Health Records (EHR) provide secure, immediate access to patient records, store  
106 patient's clinical information and coding/billing data, and allow patients secure access to their  
107 own records, thereby providing comprehensive information at the point-of-care. Following the  
108 2009 Health Information Technology for Economic and Clinical Health (HITECH) Act, EHRs  
109 have gained widespread penetration into the U.S. healthcare system, with implementation  
110 reaching 96% of U.S. hospitals by 2017 [1, 2].

111 However, for the providers, EHR usage is labor-intensive by its very nature, with its repetitive  
112 data entry now the major contributor to provider burnout nationally [3-8]. In outpatient clinics,  
113 providers typically devote ~50% of their working hours to the EHR, equating to nearly two hours  
114 of EHR use for every hour of patient consultation [9, 10]. This substantial time and effort  
115 required for EHR data entry detracts from the actual time providers spend interacting with the  
116 patient (versus interacting with the computer) during the consultation. This emphasis on EHR  
117 data entry results in suboptimal direct communication between providers and patients, detracts  
118 from the quality of the consultation, and decreases patient/provider satisfaction, contributing to  
119 the current high rates of physician burnout [11]. Additionally, providers often spend considerable  
120 amount of personal time after-hours and during weekends to complete EHR-related tasks,  
121 termed 'work-outside-work' (WOW). Small wonder then that currently up to 70% of U.S.  
122 physicians report burnout, directly attributed to EHR usage [11-14]

123 Provider burnout is linked to various aspects of EHR use, including time spent on EHR, level of  
124 organizational support, and system usability [15]. Burnout can decrease the provider's quality of  
125 EHR usage, contributing to diagnostic errors, with implications for patient care [16].

126 Misalignment of providers' cognitive processes with the technical aspects of EHR usability and  
127 workflow can increase the risk for diagnostic errors [17]. Collectively, these findings highlight the  
128 multi-dimensional burdens confronting providers when incorporating EHR into their practice.

129 During outpatient consultations, providers are required to create a Clinical Consultation Note  
130 (CCN) for each patient, which consolidates the patient's medical history, prior test data, and  
131 treatment plans in a concise and structured format. Constructing the CCN in the outpatient clinic  
132 takes on average 16 minutes of EHR usage per patient [12]. Thoroughness of the CCN drives  
133 ICD clinical coding [18], and high CCN quality is also associated with improved physician  
134 performance and patient outcomes [19]. Objective, standardized tools to assess CCN diagnostic  
135 efficiency and quality are available [20, 21], and EHR metadata can help evaluate the individual  
136 physician effort invested into creating this document [9, 10].

137 Emerging artificial intelligence (AI) and machine learning (ML) technologies present robust new  
138 opportunities for healthcare innovation [22]. Given that a significant portion of CCN creation  
139 involves repetitive, scribe-like work, it can be mitigated with technology. Efforts are underway to  
140 use speech-to-text technology and AI to automate CCN creation during clinic visits [9, 10]. More  
141 recently, a novel AI/ML and logic-based technology platform has been developed that automates  
142 the construction of a detailed preliminary CCN before the actual consultation, without any work  
143 on the part of the provider.

144 Herein, we intend to perform a prospective randomized controlled trial (RCT) to compare this  
145 novel technology platform with current standard-of-care practices in the outpatient clinic setting.  
146 We seek to evaluate the benefits and efficiencies that may accrue to providers and patients  
147 alike from using AI to automate the creation of the CCN. Using a large multi-institutional, multi-  
148 disciplinary, multi-EHR cohort, our RCT will compare the novel technology (wherein the CCN is  
149 constructed automatically using AI before the actual consultation) vis-s-vis the current standard-  
150 of-care practice (wherein the CCN is constructed manually during the actual consultation). The  
151 generated data will assess CCN diagnostic efficiency and quality, provider 'work-outside-work'  
152 effort, provider productivity, provider satisfaction, and patient satisfaction.

153

154

155 **NOVEL TECHNOLOGY**

156 The novel AI technology is comprised of a proprietary logic algorithm-based, symptom-tree  
157 Active Sheet that extracts subjective patient data directly from the patient and analyzes and  
158 reformats it in a structured manner. In parallel, the AI/ML algorithm extracts relevant,  
159 unstructured, objective data from outside records/documents. Proprietary, embedded AI/ML  
160 technology generates national guidelines-based recommendations for diagnostic testing in a  
161 patient-specific manner. This preliminary CCN is automatically presented to the provider, in  
162 advance of the patient's visit, either within the institution's EHR or in a secure web-based App, in  
163 a HIPAA-compliant manner. During the actual consultation, the provider re-confirms the clinical  
164 history with the patient, performs the physical exam, and finalizes the CCN, completing the  
165 patient consultation. The platform's combination of AI/ML algorithms, Active Sheet logic  
166 algorithms, and Optical Character Recognition (OCR) technology makes it adaptable to various  
167 types of data inputs. Its capability of EHR-integrated and web-based deployment, as well as  
168 usage via laptops and cell phones, also ensures high scalability, allowing implementation in an  
169 institution with or without an existing EHR system, and for patients with or without laptop  
170 computers.

171

172 **METHODS**

173 Our study protocol follows the Standard Protocol Items Recommendations for Interventional  
174 Trials-Artificial Intelligence (SPIRIT-AI) Guidelines [23]. The current manuscript is only a  
175 proposal for the protocol for this study; IRB approval and clinicaltrials.gov notification will  
176 precede the actual trial initiation and consenting/enrollment of any clinical patients.

177

178 **Study Settings**

179 This is a prospective multi-institutional, multi-disciplinary, multi-EHR randomized trial. To assess  
180 generalizability across the health-care landscape, the RCT will be conducted concomitantly

181 across four different types of healthcare settings: large academic medical center, non-academic  
182 hospital, rural hospital, and community private practices. Participating centers include the  
183 University of Southern California and Fox Chase Cancer Center (large academic medical  
184 centers); Sierra View Medical Center, Verdugo Hills Hospital (non-academic, rural hospitals);  
185 and Solaris Urology and Glendale Cardiology (community private practices). Four specialties  
186 will be assessed, including two medical specialties (cardiology, infectious disease), a surgical  
187 specialty (urology), and an urgent care specialty (emergency medicine). Four different types of  
188 EHRs will be assessed, including Cerner, Epic, AllScripts, and MediTech.

189 This prospective trial will randomize the novel AI/ML technology (NAMT) versus the current  
190 standard-of-care (SOC) in the outpatient setting. Randomization will be done per outpatient  
191 clinic-days based on method of constructing the CCN. As such, randomization to either the  
192 “SOC clinic-day” or the “NAMT clinic-day”, will be based on whether the SOC or the NAMT was  
193 used to construct the CCN for all patients seen by that provider in that clinic on that day.

194 Randomized cross-over of each provider between “SOC clinic-day” and “NAMT clinic-day” will  
195 allow each provider to serve as his/her own internal control. To provide a balanced and valid  
196 comparison, we will use similar clinic workflow methodology for collecting outside records/data  
197 in each arm. Thus, medical records clerks and/or nurses would collect any available outside  
198 patient records/data in both arms per existing, legacy data collection methods at each  
199 participating site. In the “SOC clinic-day” arm, the medical records clerk/nurse would upload  
200 patient records to the native EHR platform, and then the CCN would be constructed manually  
201 per current standard practice. In the “NAMT clinic-day” arm, the medical records clerk/nurse  
202 would upload these records to the NAMT platform (typically housed within the native EHR, thus  
203 maintaining similarity in data/records collection methods), but the CCN would be constructed  
204 automatically by the NAMT platform. Thus, we will use the same data entry mechanism in each  
205 arm to ensure consistency and enhance validity of results. Similar objective endpoint data will

206 be obtained for each provider and patient in the control “SOC clinic-day” arm and the  
207 experimental “NAMT clinic-day” arm (Figure 1 – CONSORT-AI Flowchart [24]).  
208 Study endpoints will compare the SOC vs the NAMT based on objective data. Co-primary  
209 endpoints are: a) CCN quality and diagnostic accuracy (based on standardized QNOTE metrics  
210 [14]); and b) Work-outside-work (WOW) time required by providers to complete clinic-related  
211 documentation tasks outside normal clinic hours (based on EHR meta-data). Secondary  
212 endpoints include: a) Provider satisfaction (based on standardized AHRQ EHR End User  
213 Survey); b) Patient satisfaction (based on standardized Press Ganey/CG-CAHPS survey  
214 ([www.pressganey.com](http://www.pressganey.com))); and c) Provider productivity (based on provider “walk-in, walk-out’ time  
215 from the individual consultation room) and imputed financial data. In total, 15-30 providers will  
216 participate. Provider is defined as a licensed medical provider who is constructing the CCN in  
217 their daily practice currently, including practicing physicians, mid-level practitioners (physician  
218 assistant, nurse practitioner), or trainees (post-graduate fellow or resident). Scribes are not  
219 designated as a provider. At the time of scheduling the clinic appointment, patients will be  
220 informed that, on their specific clinic-day, their provider is doing all consultations using the  
221 specific methodology (SOC or NAMT), as dictated by the randomization. At study onset, all  
222 participating providers will attest that they have equipoise between the SOC and NAMT  
223 approaches.

224

### 225 **Randomization: Rationale**

226 Our objective is to evaluate the merits/demerits of the novel AI/ML technology for the automated  
227 construction of CCN vis-a-vis the standard-of-care (SOC). Elements that could potentially be  
228 randomized include patients, providers, or the intervention itself. However, practically speaking,  
229 all of these would impact daily clinic workflow. For instance, if we were to randomize patients,  
230 then using both the SOC and NAMT within the same clinic day would be highly disruptive to  
231 clinic workflow, and challenging, if not prohibitive, to implement from a practical perspective. If



232 we randomized providers, it would be problematic to randomize the AI/ML technology, and vice-  
233 versa. Also, providers with different practice styles conducting one or the other intervention  
234 would not control for inter-personal disparities. However, randomization of the entire clinic-day  
235 to either “SOC clinic-day” or “NAMT clinic-day” for all patients seen that day, and then equally  
236 alternating each provider between “SOC clinic-day” and “NAMT clinic-day” would satisfy study  
237 objectives and allow each provider to serve as her/his own internal control. Providers would be  
238 a heterogeneous group as regards gender, race, and age. This study design is pragmatic and  
239 would be minimally disruptive to daily clinic workflow or provider productivity. To minimize  
240 provider bias, at enrollment, all consented providers would attest to equipoise between SOC  
241 and NAMT.

242

### 243 **Eligibility Criteria**

244 Inclusion and exclusion criteria are reported separately for providers and patients.

245

246 a) Providers:

247 i. Inclusion criteria: Licensed medical providers currently engaged in patient  
248 consultations in the outpatient clinic, including practicing physicians; mid-level  
249 practitioners (physician assistant, nurse practitioner); or trainee (post-graduate fellow  
250 or resident). Use of scribes by providers is allowed.

251 ii. Exclusion Criteria: Licensed medical providers who are currently not engaged in  
252 patient consultations in the outpatient clinic.

253

254 b) Patients:

255 i. Inclusion criteria: Outpatient clinic patients at least 18 years of age or older being  
256 seen in the cardiology, infectious disease, urology, and emergency medicine  
257 outpatient settings.

258 ii. Exclusion criteria: Patients younger than 18 years of age.

259

## 260 **Intervention**

261 Licensed, actively practicing providers who satisfy inclusion criteria are recruited after written  
262 consent. Their upcoming clinic days are randomized to either “SOC clinic-day” or “NAMT clinic-  
263 day”, wherein CCNs for all patients seen on those respective days are constructed using either  
264 the standard-of-care practice or the novel AI/ML technology (Figure 1). Clinic days will be  
265 randomized 1-2 weeks prior to the actual date of that clinic. As such, all patients seen on the  
266 clinic day randomized to the “SOC clinic-day” will have their CCN constructed per that provider’s  
267 customary, standard-of-care practice. Patients seen on the clinic day randomized to the “NAMT  
268 clinic-day” will have their CCN constructed 2 weeks to 1 day prior to the clinic visit using the  
269 novel AI/ML technology. The automatically constructed CCN will be made available to the  
270 provider in advance of the clinic day, either integrated within their native EHR or in the secure,  
271 free-standing web-based App. After pre-viewing the CCN, the provider will proceed to perform  
272 the clinical consultation. Provider “walk-in, walk-out” time from the consultation room will be  
273 recorded by a third-party observer using a stopwatch. After completing the consultation, the  
274 patient will be invited to complete the adapted Press Ganey/CG-CAHPS patient satisfaction  
275 questionnaire; the patient will provide oral or written informed consent before filling out the de-  
276 identified survey instrument. This survey questionnaire has been adapted to fit our study-related  
277 objectives, including questions on patient age, gender, level of education, race/ethnicity,  
278 provider attentiveness to the patient, provider spending time doing computer work versus  
279 interacting with the patient, and provider’s prior knowledge about the patient’s clinical condition  
280 before entering the patient’s consultation room. The questionnaire then provides the patient a  
281 final opportunity to record an overarching ‘thumbs-up’ or ‘thumbs-down’ about their clinic  
282 experience. Interventions for each group are provided below with sufficient details to allow  
283 replication, including method and timing of administration.

284

285 **a) Study Intervention for providers**

286 Providers will be invited to participate in the study and informed consent obtained. Providers  
287 and their teams will be trained on the novel technology and related clinical workflow over a  
288 period of 1-3 clinic-days (“Wash-in” period) to make them knowledgeable about the novel  
289 technology platform.

290 Outpatient clinic days will be randomized into one of 2 arms, “SOC clinic-day” and “NAMT clinic-  
291 day”. (Details in Figure 1)

292

- 293 i. On “SOC clinic-day”: CCNs for all patients seen on this day will be constructed per  
294 the usual, standard-of-care practice of that provider. Prior to the clinic day, the  
295 medical records clerk/nurse would upload any outside patient records/data to the  
296 native EHR platform per current standard practice, and then the CCN would be  
297 constructed manually.
- 298 ii. On “NAMT clinic-day”: CCNs for all patients seen on this day will be pre-constructed  
299 using the novel AI/ML technology. At the time of being given the clinic appointment,  
300 patients will be sent the electronic link and advised to do two things from their home -  
301 ‘the earlier, the better’ - prior to coming to the clinic: a) complete the Active Sheet  
302 (AS): when completing the AS, the patient may solicit help from family/friends, or  
303 from remote licensed provider extender (PA/NP) or referring provider staff; and b)  
304 upload any lab/radiology test results: the medical records clerk/nurse would help the  
305 patient upload outside records/data to the NAMT platform (typically housed within the  
306 native EHR platform). After these two actions – completing the AS; uploading the  
307 lab/radiology test results - are done, the novel AI software automatically constructs  
308 the CCN and delivers it to the provider, either in their native EHR or in the secure,  
309 free-standing web-app. This action occurs automatically in the background invisible

310 to the provider/staff, in advance of the patient's visit. The patient then comes to the  
311 clinic on the scheduled appointment date, the provider reviews the pre-constructed  
312 CCN and proceeds with the consultation.

313

#### 314 **b) Study Intervention for patients**

315 Immediately upon completion of their consultation, patients will be invited to fill out the adapted  
316 Press-Ganey/CG-CAHPS survey in a de-identified fashion to assess patient satisfaction.  
317 Patients who agree to fill out the survey will be asked to first provide written informed consent  
318 for inclusion in the study. Upon completion of the survey, the patient departs the clinic area.  
319 Patient participation will be deemed completed upon survey completion.

320

#### 321 **Outcomes and Measurements**

322 Table 1 summarizes the a) co-primary and b) secondary outcomes of interest. Details regarding  
323 the outcome measurements are reported in the supplementary materials.

324

#### 325 **a) Co-Primary Outcomes**

326 Our co-primary objectives are (a) delivering a CCN of high-quality and diagnostic accuracy  
327 without any effort on the part of the provider, and (b) reducing the provider time for doing work-  
328 outside-work (WOW) to complete clinic-related documentation activities performed outside of  
329 the typical 8 am to 4 pm clinic schedule. Corresponding co-primary endpoints are the objective  
330 data metrics to assess CCN quality and diagnostic accuracy (using the standardized QNOTE  
331 methodology), and provider WOW time (using EHR meta-data). The rationale behind these  
332 endpoints is rooted in the central tenet of this automated technology, which is to create an  
333 automated CCN that matches or exceeds the current, manually created, standard-of-care CCN,  
334 while minimizing workload for the healthcare provider. Furthermore, we aspire to decrease the  
335 time spent by providers working on clinic notes outside regular work hours. To ascertain these

336 endpoints, we utilize QNOTE to measure CCN quality and diagnostic accuracy using objective  
337 metrics; provider's WOW time is determined by analyzing the objective EHR metadata for that  
338 clinic-day, provided by the institutional IT team. The putative mechanisms of action underpinning  
339 these objectives lie in streamlining clinical processes, thus enhancing the quality of patient care  
340 and reducing diagnostic errors.

341

## 342 **b) Secondary Outcomes**

343 The secondary objectives of the study assess provider satisfaction, patient satisfaction, and  
344 physician productivity. Provider satisfaction will be assessed using an adapted AHRQ Electronic  
345 Health Record End User Survey. We hypothesize that automating the CCN will reduce the  
346 workload of physicians, and consequently improve satisfaction and decrease burnout rates. At a  
347 minimum at the end of each clinic day, providers will fill out the AHRQ EHR End User Survey to  
348 capture their clinic-day experiences and responses.

349 Patient satisfaction, another critical component of healthcare delivery, will be evaluated using an  
350 adapted Press Ganey/CG-CAHPS survey. We hypothesize that if providers spend more time  
351 directly interacting with patients, instead of interacting with and entering data into the EHR,  
352 patients will experience higher satisfaction. As such, consented patients will be requested to  
353 complete the adapted Press Ganey/CG-CAHPS survey immediately after their consultation to  
354 measure their level of satisfaction. Finally, physician productivity, particularly in terms of time  
355 savings, which can lead to increased patient throughput and financial return-on-investment  
356 (ROI), will be documented. We believe that automating the CCN might lead to decreased, yet  
357 higher quality, consultation times, thereby improving physician productivity. To measure this, the  
358 physician's "walk-in, walk-out" time from each consult room will be recorded with a stopwatch.  
359 This objective, real-time measurement will assess the impact of the intervention on physician  
360 productivity. Financial ROI will be imputed.

361

## 362 **Sample Size**

363 Given the novelty of our AI algorithm, prior randomized data do not exist. Our sample size will  
364 be designed to detect a minimum of 30% improvement in the CCN quality/diagnostic accuracy,  
365 and a 30% reduction in the duration of WOW, our 2 co-primary endpoints. We estimate the total  
366 sample size to be 150 clinic-days, i.e., 75 “SOC-days” and 75 “NAMT-days”. Conservatively  
367 assuming 20 patients seen per clinic-day, we will accrue at least 3,000 total patients - 1,500 in  
368 the control “SOC-day” arm, and 1,500 in the intervention “NAMT-day” arm. Clinic-days will be  
369 divided between the 4 specialties, leading to 30-40 clinic-days per specialty. We estimate a total  
370 of 15-30 study providers, resulting in 5-10 clinic-days per provider. No single specialty will  
371 accrue more than 30% of the aggregate clinic-days. We will use a two-sided Z-test of the  
372 difference between proportions with 90% power and two-sided 5% significance level. This  
373 sample size exceeds minimum requirements and will provide robust data. Study data will be  
374 collected by trained personnel, physically located at each clinical site. The validated QNOTE  
375 scoring system will assesses quality and diagnostic accuracy of de-identified CCNs uploaded to  
376 an online repository for scoring by independent, blinded raters. Data collection process will be  
377 integrated into clinic workflows, preserving the user experience while introducing the new AI/ML  
378 tool.

379

## 380 **Recruitment**

381 Providers will be recruited from the participating institutions. We anticipate between 15-30  
382 providers to participate in this RCT. This would provide a heterogeneous group of providers as  
383 regards gender, race, and age, with each provider participating in 5-10 study clinic-days.  
384 Randomized alternating cross-over of providers will result in each provider serving as her/his  
385 own internal control. Patients will be identified exclusively based on the participating providers'  
386 clinic schedules. Patients will be recruited and consented to the study immediately after they  
387 finish their clinical consultation with the provider, and before they are administered the Press

388 Ganey/CG-CAHPS survey instrument. We intend to recruit over 3000 patients over a one-year  
389 period. No incentives will be provided to either physicians or patients for participating in this  
390 RCT.

391

### 392 **Screening Failure**

393 In case a provider has been already enrolled but wishes to drop out of the study within the first 5  
394 clinic-days, he/she will be replaced with another physician from that specialty. Any provider who  
395 has already completed the first 5 clinic-days in the study will not be replaced by another  
396 provider; their data will be used as is with censoring at study drop-out. If a patient chooses to  
397 not fill out the Press Ganey/CG-CAHPS satisfaction survey, it will be recorded as such, and their  
398 survey data will be excluded from study calculations.

399

### 400 **Discontinuation of intervention**

401 After initially agreeing to participate, if a provider decides to discontinue in the study, she/he  
402 must express this desire prior to 5 clinic-days after enrolling in the study; in this case their data  
403 over those 5 clinic-days will be used, and they will be replaced by another physician. After 5  
404 clinic-days, they will be considered to have been part of the trial and while their data will be  
405 included for analyses, they will not be replaced by another physician. Participation in this study  
406 is voluntary. If a patient decides to not provide informed consent at the time of answering the  
407 Press Ganey/CG-CAHPS questionnaire, their survey data will not be included in the study. If a  
408 patient decides to not participate in the study after initially agreeing, they must express this  
409 desire prior to completing their Press Ganey/CG-CAHPS questionnaire and leaving the clinic.  
410 The sole criterion for patient participation in the study is patient willingness. Participation in this  
411 study is voluntary.

412

### 413 **Statistical Hypothesis and Analysis**

414

415 a) Statistical Hypothesis

416 i. Primary Efficacy Endpoint(s): We hypothesize that the CCN generated by the novel  
417 technology will be superior to the SOC as regards CCN quality and diagnostic  
418 accuracy (as measured by QNOTE methodology), and CCN efficiency (WOW time  
419 as measured by EHR meta-data). Our null hypothesis is that there will be no  
420 difference in the quality, diagnostic accuracy, and efficiency of the CCN derived from  
421 the novel technology compared to standard-of-care.

422 ii. Secondary Efficacy Endpoint(s): We hypothesize that the CCN generated by the  
423 novel AI/ML technology will be superior to the SOC as regards provider satisfaction  
424 (adapted AHRQ Electronic Health Record End User Survey); patient satisfaction  
425 (adapted Press Ganey/CG-CAHPS survey); and physician productivity (“walk-in,  
426 walk-out” consult time).

427

428 Our null hypothesis is that outcomes will be no difference in the above-mentioned three metrics  
429 for CCNs generated by the novel AI/ML technology versus standard-of-care.

430

431 b) Statistical analysis

432 Analysis of the primary endpoint will assess mean data for CCN quality/accuracy and CCN  
433 efficiency to demonstrate superiority of the NT compared to SOC.

434

435 i. Quality of CCN will be assessed by the adapted QNOTE instrument, a validated  
436 methodology that utilizes 15 clinical elements and 7 evaluative components (as  
437 detailed in supplementary materials). Each CCN will be de-identified by redacting  
438 patient/physician identifiers prior to assessment and uploaded to an online repository  
439 for QNOTE scoring by three blinded, third-party raters. Each patient will be assigned



440 a unique identifier number to ensure confidentiality and privacy. Depending on quality  
441 of each CCN, a score between 0 (minimum) to 1500 (maximum) will be assigned by  
442 independent raters. The average of the ratings from the 3 independent raters for  
443 each CCN will denote the composite QNOTE score for that CCN, which will be the  
444 basis for statistical analysis comparing CCNs created on “SOC clinic-day” versus  
445 “NT clinic-day”.

446 ii. Work-outside-Work (WOW) time will be objectively assessed from EHR meta-data  
447 per provider (as detailed in supplementary materials).

448

449 Analysis of the secondary endpoints will calculate mean data to assess outcomes of the NT  
450 compared to SOC for the following three aspects:

451

452 i. Provider satisfaction (adapted AHRQ Electronic Health Record End User Survey):  
453 dichotomous positive versus negative Likert scale answers will be compared for each  
454 question and in the aggregate.

455 ii. Patient satisfaction (adapted Press Ganey/CG-CAHPS survey): dichotomous  
456 positive versus negative Likert scale answers will be compared for each question  
457 and in the aggregate.

458 iii. Provider productivity (Total consultation time): physician “walk in-walk out” time from  
459 the consult room at each patient consultation will be measured by the study  
460 coordinator using a stopwatch.

461

462 The sample size will be designed to detect a minimum of 30% improvement in the CCN  
463 quality/diagnostic accuracy, and a 30% reduction in the duration of WOW, our 2 co-primary  
464 endpoints. Analysis of primary and secondary objectives will proceed as follows. Missing data  
465 will be dealt with by imputation, and potential outliers will be scrutinized by the Principal

466 Investigator. For each question, Likert scale responses will be analyzed, allowing comparisons  
467 between the novel technology and the standard-of-care. Answers will be divided into positive  
468 and negative categories per the "top box method," which is standard for assessing CG-CAHPS  
469 data. Likert scale responses will be dichotomized, with scores of 4 (agree) and 5 (strongly  
470 agree) representing agreement (positive) and scores of 1 (strongly disagree), 2 (disagree), or 3  
471 (neutral) indicating disagreement (negative). Screening for outliers will be conducted based on  
472 the absolute agreement of each individual question and distribution of responses. Continuous  
473 variables will be represented as mean and standard deviation (SD), and categorical variables  
474 will be presented as median and interquartile range (IQR). For the univariate analysis, Kruskal-  
475 Wallis, chi-squared ( $\chi^2$ ), and Fisher's exact tests will be utilized to compare continuous and  
476 categorical variables as appropriate. A 95% confidence interval and a two-tailed test with  $p < .05$   
477 will be taken as statistically significant.

478 Upon completion of data collection, each question will be analyzed using Cochran-Mantel-  
479 Haenszel tests, stratified by provider to eliminate potential individual provider bias from the  
480 results. P-values will be determined using Chi-Square tests and will be deemed significant at  
481 less than 0.05. For the assessment of patient satisfaction (utilizing the adapted Press  
482 Ganey/CG-CAHPS survey with additional queries), each question will be stratified by provider to  
483 remove potential individual bias and analyzed using Cochran-Mantel-Haenszel tests.

484 Considering the study's novelty and the absence of prior randomized data, an interim analysis  
485 will be conducted. Calculations will be performed using an  $\alpha$  level of 0.05 and a power of 90%.  
486 Based on the trends in the data from the first 50 patients, sample sizes will be calculated for  
487 each question, assuming the variables to be dichotomous. After the complete data collection,  
488 each question will be analyzed using Cochran-Mantel-Haenszel tests, stratified by providers to  
489 remove any individual provider bias. P-values will be determined using Chi-Square tests, and a  
490 value of less than 0.05 will be considered significant.

491

492 **CONCLUSION**

493 Herein we propose the protocol of the first prospective randomized controlled trial of a large-  
494 scale, multi-institutional, multi-disciplinary cohort to evaluate the practicality and efficiency of  
495 using AI/ML technology to pre-construct the CCN automatically before the actual patient-  
496 provider consultation. The study will compare this AI/ML framework vis-a-vis the standard-of-  
497 care, wherein the CCN is generated manually during the consultation encounter. Data  
498 generated by this study will provide insights into the desirability and generalizability of AI-  
499 generated automated clinical consultation notes across the healthcare spectrum, and its  
500 potential benefits to providers, patients, and healthcare institutions.

501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533

534 **REFERENCES**

535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584

1. Goldstein, M.M. and H. Thorpe Jane, The First Anniversary of the Health Information Technology for Economic and Clinical Health (HITECH) Act: the regulatory outlook for implementation. *Perspectives in health information management*, 2010. 7(Summer): p. 1c.
2. HealthIT.gov, <https://www.healthit.gov/data/quickstats/office-based-physician-electronic-health-record-adoption>. 2020.
3. McPeck-Hinz, E., et al., Clinician Burnout Associated With Sex, Clinician Type, Work Culture, and Use of Electronic Health Records. *JAMA Netw Open*, 2021. 4(4): p. e215686.
4. Melnick, E.R., et al., Analysis of Electronic Health Record Use and Clinical Productivity and Their Association With Physician Turnover. *JAMA Netw Open*, 2021. 4(10): p. e2128790.
5. Rotenstein, L.S., C. Sinsky, and C.K. Cassel, How to Measure Progress in Addressing Physician Well-being: Beyond Burnout. *Jama*, 2021.
6. Kroth, P.J., et al., Association of Electronic Health Record Design and Use Factors With Clinician Stress and Burnout. *JAMA Netw Open*, 2019. 2(8): p. e199609.
7. Melnick, E.R., et al., The Association Between Perceived Electronic Health Record Usability and Professional Burnout Among US Physicians. *Mayo Clin Proc*, 2020. 95(3): p. 476-487.
8. Nguyen, O.T., et al., A systematic review of contributing factors of and solutions to electronic health record-related impacts on physician well-being. *J Am Med Inform Assoc*, 2021. 28(5): p. 974-984.
9. Kumah-Crystal, Y.A., et al., Electronic Health Record Interactions through Voice: A Review. *Applied clinical informatics*, 2018. 9(3): p. 541-552.
10. Kataria, S. and V. Ravindran, Electronic health records: a critical appraisal of strengths and limitations. *J R Coll Physicians Edinb*, 2020. 50(3): p. 262-268.
11. Naidu, A., Factors affecting patient satisfaction and healthcare quality. *Int J Health Care Qual Assur*, 2009. 22(4): p. 366-81.
12. Overhage, J.M. and D. McCallie, Physician Time Spent Using the Electronic Health Record During Outpatient Encounters. *Ann Intern Med*, 2020. 173(7): p. 594-595.
13. Yan, Q., et al., Exploring the relationship between electronic health records and provider burnout: A systematic review. *J Am Med Inform Assoc*, 2021. 28(5): p. 1009-1021.
14. Burke, H.B., et al., QNOTE: an instrument for measuring the quality of EHR clinical notes. *J Am Med Inform Assoc*, 2014. 21(5): p. 910-6.
15. Li, C., et al., Electronic medical record-related burnout in healthcare providers: a scoping

585  
586 review of outcomes and interventions. *BMJ Open*, 2022. 12(8): p. e060865.  
587  
588 16. Graber, M.L., C. Byrne, and D. Johnston, The impact of electronic health records on  
589 diagnosis. *Diagnosis (Berl)*, 2017. 4(4): p. 211-223.  
590  
591 17. Dixit, R.A., et al., Electronic Health Record Use Issues and Diagnostic Error: A Scoping  
592 Review and Framework. *J Patient Saf*, 2023. 19(1): p. e25-e30.  
593  
594  
595 18. (2008)., A.C.f.L.A.a.t.N.H.S.i.E.a.W., PbR Data Assurance Framework 2007/08: findings  
596 from the first year of the national clinical coding audit programme. Audit Commission. 2008.  
597  
598 19. Mann, R. and J. Williams, Standards in medical record keeping. *Clin Med (Lond)*, 2003.  
599  
600 3(4): p. 329-32.  
601  
602 20. Logan, J.R., P.N. Gorman, and B. Middleton, Measuring the quality of medical records: a  
603 method for comparing completeness and correctness of clinical encounter data. *Proc AMIA*  
604 *Symp*, 2001: p. 408-12.  
605  
606 21. Hong, C.J., et al., Accuracy and completeness of electronic medical records obtained  
607 from referring physicians in a Hamilton, Ontario, plastic surgery practice: A prospective  
608 feasibility study. *Plast Surg (Oakv)*, 2015. 23(1): p. 48-50.  
609  
610 22. Lee, P., S. Bubeck, and J. Petro, Benefits, Limits, and Risks of GPT-4 as an AI Chatbot  
611 for Medicine. *N Engl J Med*, 2023. 388(13): p. 1233-1239.  
612  
613 23. Cruz Rivera, S., et al., Guidelines for clinical trial protocols for interventions involving  
614 artificial intelligence: the SPIRIT-AI extension. *Lancet Digit Health*, 2020. 2(10): p. e549-e560.  
615  
616 24. Liu, X., et al., Reporting guidelines for clinical trial reports for interventions involving  
617 artificial intelligence: the CONSORT-AI extension. *Lancet Digit Health*, 2020. 2(10): p. e537-  
618 e548.  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635

636 **Table 1**  
637

OBJECTIVES	ENDPOINTS	JUSTIFICATION FOR ENDPOINTS	PUTATIVE MECHANISMS OF ACTION
<b>CO-PRIMARY</b>			
<p>1. Quality/Diagnostic Accuracy of Clinical Consultation Note (CCN)</p> <p>2. Efficiency of CCN construction</p>	<p>1. QNOTE: to assess CCN quality/diagnostic accuracy (<i>Objective data</i>)</p> <p>2. Objective Meta-data: from EMR (<i>Objective data</i>)</p>	<p>1. The quality/accuracy of the automated CCN should be, at a minimum, equivalent to the standard-of-care</p> <p>2. Extra time (beyond typical clinic hours) spent by provider to complete CCN should decrease</p>	<p>1. QNOTE: Blinded, objective third-party assessment of CCN quality/accuracy</p> <p>2. Work-outside-Work (WOW) time/provider: Objective per-patient EHR meta-data from IT team</p>
<b>SECONDARY</b>			
<p>3. Provider satisfaction</p> <p>4. Patient satisfaction</p> <p>5. Provider productivity</p>	<p>3. AHRQ EHR End User Survey (<i>Objective data</i>)</p> <p>4. Press Ganey/CG-CAHPS survey (<i>Objective data</i>)</p> <p>5 Consultation time savings (<i>Objective data</i>)</p>	<p>3. Automating the CCN should decrease provider workload &amp; improve satisfaction</p> <p>4. Automating the CCN should allow provider to spend more time interacting with the patient (rather than with the EHR), improving patient satisfaction</p> <p>5. Consultation times should, at a minimum, be similar to standard-of-care, upon automating the CCN</p>	<p>3. Providers to fill out adapted AHRQ EHR End User Survey per new patient consult or per clinic-day</p> <p>4. Patients to fill out adapted Press Ganey/CG-CAHPS survey at end of the consult</p> <p>5. Physician “walk-in walk-out” time from consult room, measured by stopwatch</p>

638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658

### CONSORT Flow Diagram (Ideal Scenario)

