

# **Large language models can match junior clinicians in discharge letter writing: a single-blinded study**

Joshua Yi Min Tung, Sunil Ravinder Gill, Gerald Gui Ren Sng, Daniel Yan Zheng Lim, Yuhe Ke, Ting Fang Tan, Liyuan Jin, Kabilan Elangovan, Jasmine Chiat Ling Ong, Hairil Rizal Abdullah, Daniel Shu Wei Ting, Tsung Wen Chong

Submitted to: Journal of Medical Internet Research  
on: February 25, 2024

**Disclaimer:** © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

*Table of Contents*

Original Manuscript ..... 5

# Large language models can match junior clinicians in discharge letter writing: a single-blinded study

Joshua Yi Min Tung<sup>1,2</sup> MBBS, MPH; Sunil Ravinder Gill<sup>1</sup> MBBS; Gerald Gui Ren Sng<sup>2,3</sup>; Daniel Yan Zheng Lim<sup>2,4</sup>; Yuhe Ke<sup>2,5</sup>; Ting Fang Tan<sup>6</sup>; Liyuan Jin<sup>7</sup>; Kabilan Elangovan<sup>8</sup>; Jasmine Chiat Ling Ong<sup>9</sup>; Hairil Rizal Abdullah<sup>2,5</sup>; Daniel Shu Wei Ting<sup>6,7,8</sup>; Tsung Wen Chong<sup>1</sup> DPhil

<sup>1</sup>Department of Urology Singapore General Hospital Singapore SG

<sup>2</sup>Data Science and Artificial Intelligence Lab Singapore General Hospital Singapore SG

<sup>3</sup>Department of Endocrinology Singapore General Hospital Singapore SG

<sup>4</sup>Department of Gastroenterology and Hepatology Singapore General Hospital Singapore SG

<sup>5</sup>Department of Anesthesiology Singapore General Hospital Singapore SG

<sup>6</sup>Singapore Eye Research Institute Singapore National Eye Center Singapore SG

<sup>7</sup>Duke-NUS Medical School Singapore SG

<sup>8</sup>Artificial Intelligence Office Singapore Health Services Singapore SG

<sup>9</sup>Division of Pharmacy Singapore General Hospital Singapore SG

## Corresponding Author:

Joshua Yi Min Tung MBBS, MPH

Department of Urology

Singapore General Hospital

Outram Road

Singapore

SG

## Abstract

**Background:** Discharge letters are a critical component in continuity of care between specialists and primary care providers, but are time-consuming to write, under-prioritized in comparison to direct clinical care, and are often tasked to junior doctors. Prior studies assessing the quality of discharge summaries written for inpatient hospital admissions show inadequacies in many domains. Large language models such as GPT have the ability to summarize large volumes of unstructured free text, such as electronic medical records, and have the potential to automate such tasks, providing time savings and consistency in quality.

**Objective:** To assess the performance of GPT-4 in generating discharge letters written from Urology specialist outpatient clinics to primary care providers, and compare their quality against letters written by junior clinicians.

**Methods:** Fictional electronic records were written by physicians, simulating five common Urology outpatient cases with long-term follow-up. Records comprised simulated consultation notes, referral letters and replies, and relevant discharge summaries from inpatient admissions. GPT-4 was tasked to write discharge letters for these cases, with a specified target audience of primary care providers who would be continuing the patient's care. Prompts were written for safety, content, and style. Concurrently, junior clinicians were provided with the same case records and instructional prompts. GPT-4 output was assessed by the study team for instances of hallucination. A blinded panel of primary care physicians then evaluated the letters using a standardized questionnaire tool.

**Results:** GPT-4 outperformed human counterparts in information provision, but was less concise. GPT-4 had no instances of hallucination. There were no statistical differences in the clarity, collegiality, follow-up recommendations, and overall satisfaction between letters generated by humans and by GPT-4.

**Conclusions:** Discharge letters written by GPT-4 had equivalent quality to those written by junior clinicians, without any hallucinations. This study demonstrates proof of concept that LLMs can be useful and safe tools in clinical documentation.

(JMIR Preprints 25/02/2024:57721)

DOI: <https://doi.org/10.2196/preprints.57721>

## Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible to all users.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <http://www.jmir.org/preprint/57721>

## Original Manuscript

## Large language models can match junior clinicians in discharge letter writing: a single-blinded study

Tung Joshua Yi Min, MPH<sup>1, 2</sup>

Gill Sunil Ravinder, MBBS<sup>1</sup>

Sng Gerald GR, MPH<sup>2,3</sup>,

Lim Daniel Yan Zheng, MTech<sup>2,4,5</sup>,

Ke Yuhe, MBBS<sup>6</sup>

Tan Ting Fang, MBBS<sup>7</sup>

Jin Liyuan, B.Eng.<sup>5</sup>

Kabilan Elangovan, B.Eng.<sup>8</sup>

Ong Jasmine Chiat Ling, PharmD<sup>5,9</sup>

Hairil Rizal Abdullah<sup>2,6</sup>

Ting Daniel Shu Wei<sup>5,7,8</sup>,

Chong Tsung Wen, DPhil<sup>1</sup>

<sup>1</sup>Department of Urology, Singapore General Hospital, Singapore

<sup>2</sup>Data Science and Artificial Intelligence Lab, Singapore General Hospital, Singapore

<sup>3</sup>Department of Endocrinology, Singapore General Hospital, Singapore

<sup>4</sup>Department of Gastroenterology and Hepatology, Singapore General Hospital, Singapore

<sup>5</sup>Duke-NUS Medical School

<sup>6</sup>Department of Anesthesiology, Singapore General Hospital, Singapore

<sup>7</sup>Singapore Eye Research Institute, Singapore National Eye Center

<sup>8</sup>Artificial Intelligence Office, Singapore Health Services, Singapore

<sup>9</sup>Division of Pharmacy, Singapore General Hospital, Singapore

### Corresponding Author

Dr Joshua Tung Yi Min

16 College Road, Block 4 Level 1, Urology Centre, Singapore General Hospital

Singapore 169854

+65 97766119

joshua.tung@mohh.com.sg

### Conflicts of Interest

The authors have no conflicts of interest to declare

No funding was received for this work

### Running Title

AI matches doctors in discharge letters

### Keywords

Artificial intelligence, large language models, discharge summaries, continuity of care

## Abstract

### Background

Discharge letters are a critical component in continuity of care between specialists and primary care providers, but are time-consuming to write, under-prioritized in comparison to direct clinical care, and are often tasked to junior doctors. Prior studies assessing the quality of discharge summaries written for inpatient hospital admissions show inadequacies in many domains. Large language models such as GPT have the ability to summarize large volumes of unstructured free text, such as electronic medical records, and have the potential to automate such tasks, providing time savings and consistency in quality.

### Objective

To assess the performance of GPT-4 in generating discharge letters written from Urology specialist outpatient clinics to primary care providers, and compare their quality against letters written by junior clinicians.

### Methods

Fictional electronic records were written by physicians, simulating five common Urology outpatient cases with long-term follow-up. Records comprised simulated consultation notes, referral letters and replies, and relevant discharge summaries from inpatient admissions. GPT-4 was tasked to write discharge letters for these cases, with a specified target audience of primary care providers who would be continuing the patient's care. Prompts were written for safety, content, and style. Concurrently, junior clinicians were provided with the same case records and instructional prompts. GPT-4 output was assessed by the study team for instances of hallucination. A blinded panel of primary care physicians then evaluated the letters using a standardized questionnaire tool.

### Results

GPT-4 outperformed human counterparts in information provision, but was less concise. GPT-4 had no instances of hallucination. There were no statistical differences in the clarity, collegiality, follow-up recommendations, and overall satisfaction between letters generated by humans and by GPT-4.

### Conclusion

Discharge letters written by GPT-4 had equivalent quality to those written by junior clinicians, without any hallucinations. This study demonstrates proof of concept that LLMs can be useful and safe tools in clinical documentation.

Word count: 2010

## Introduction

Large language models (LLMs) are artificial intelligence (AI) constructs with the capability to parse and generate human-like text based on pre-training with a vast corpus of unstructured text data. LLMs have shown promise in parsing large corpi of text data and generating human-like text across various domains, including healthcare<sup>1</sup>.

One of the most promising applications of LLMs in healthcare is for automating tasks traditionally performed by healthcare professionals, such as writing discharge summaries and letters<sup>2</sup>. Despite the fact that such documents serve as a critical communication tool between hospital specialists and primary care providers (PCPs) in ensuring continuity of care, they are time-consuming to write, are often tasked to junior doctors, and are under-prioritized when balanced against direct clinical care<sup>3,4</sup>.

Prior studies have shown that inpatient hospital discharge summary quality is inadequate in many domains/elements required by The Joint Commission and endorsed at the Transitions of Care Consensus Conference (TOCCC)<sup>3</sup>. Many elements, such as provision of follow-up care plans, are of particular importance to patients being discharged to primary care. Improving discharge summary timeliness was shown to have positive effects on the quality of content and transmission rate to primary care providers<sup>5,6</sup>.

The automation of inpatient discharge summaries has previously been explored, utilizing electronic discharge summary (EDS) programs. Maslove *et al* showed no differences in PCP-reported overall quality between summaries generated by an EDS program and conventional dictated summaries<sup>7</sup>. These studies assessed communication between inpatient and outpatient providers for single episodes of hospital admissions, rather than summative information regarding a patient's clinical condition across a longer follow-up, as would be the case with specialist outpatient care.

However, their applications in outpatient clinical documentation remain largely unexplored. Unlike inpatient admissions, which are typically self-contained and time-limited, outpatient specialist care can have more complex clinical documentation with longitudinal health records that may span across years. It remains to be seen if LLMs are able to parse clinical information with a strong temporal dimension and produce high-quality output.

There is currently no consensus on the single best way to objectively evaluate clinically-relevant LLM output. However, approaches comparing LLM output against that generated by clinicians can help with assessment of a benchmark of "human-like" or "clinician-like" performance, and have been previously described in the literature<sup>8</sup>. In addition, independent expert evaluation of LLM output has also been used for physician evaluation of LLM output<sup>9</sup>. We sought to combine both approaches in this study.

This study aimed firstly to evaluate the feasibility and quality of an LLM in writing discharge letters for patients discharged from a specialist outpatient clinic, and secondly to compare the discharge letters generated against similar letters written by human junior clinicians.



## Methods

This study was conducted in a simulated environment using only fictional patient data.

### Data Generation

Fictional patient data were generated by clinicians from the department of Urology, mimicking typical electronic medical records (EMRs) of 5 common patient groups seen at the outpatient clinic in a tertiary centre. The time horizons for follow-up ranged from 6 months to 6 years. Data included initial referral letters from PCPs or Emergency Departments, initial and follow-up urology consultation notes, referrals to other departments, correspondence notes, and discharge summaries from relevant admissions. Information on the fictional cases is presented in **Table 1**. Each case varied in complexity, ranging between 4-10 documents and 1238-3009 characters (424-1110 tokens).

**Table 1. Fictional case-mix data utilized for assessment**

S/N	Initial Referral	Clinical Condition	Follow-up Duration	Additional/Incidental Clinical Information
1	Elevated PSA	Benign prostatic hyperplasia	6 years	Initially declined biopsy and opted for PSA monitoring, but ultimately underwent a transperineal prostate biopsy and developed acute retention of urine post-operatively.
2	LUTS	Overactive bladder	3 years	Failed behavioural modification measures and experienced adverse effects from multiple lines of anticholinergic medications.
3	Asymptomatic microhematuria	Renal calculus	1.5 years	Incidental pancreatic lesion on imaging and was referred to General Surgery for further evaluation
4	Ureteric colic	Distal ureteric stone	6 months	Failed medical expulsive therapy and underwent ureteroscopy and lithotripsy, also noted to have hypertension peri-operatively.
5	Erectile dysfunction	Erectile dysfunction	2.5 years	Noted to have abnormal ECG and was referred to Cardiology for further assessment prior to initiation of PDE-5 inhibitors.

PSA; Prostate-specific Antigen, LUTS; Lower urinary tract symptoms, ECG; Electrocardiography, PDE-5; Phosphodiesterase-5.

### Discharge Letter Generation

GPT-4 is a state-of-the-art LLM developed by OpenAI with a parameter count of 1.76 trillion, and has exhibited human-level performance on professional and academic benchmarks<sup>10</sup>. GPT-4 with an 8K token context length was provided with the fictional EMRs and tasked to write discharge letters to PCPs. The prompt instructed GPT-4 to assume the role of a physician assistant in a Urology clinic, and specified the context of the task (“this patient is being discharged”) and the target audience (“meant to be read by the General Practitioner”). Prompts for safety included specific instructions to include only information provided in the fictional EMR, to reduce the risk of hallucination<sup>11</sup>. Desired content guidelines were provided to standardize the structure of the discharge letter. Prompts for style included instructions to write in prose, in a cordial and concise manner, and not to exceed half a page of text unless necessary. To evaluate the generative capacity of the LLM, only the above instructions (but no examples) were provided to GPT-4 - a technique also known as “zero-shot” prompting<sup>12</sup>.

Three separate discharge letters were generated for each fictional case. One was generated by GPT-4, and the comparators were two letters that were written concurrently by junior clinicians from the department of Urology using the same set of fictional patients. For the comparator letters, use of generative AI tools or other automated summarisation methods was disallowed.

Full prompt instructions are shown in **Supplementary Figure S1**. Sample medical records and discharge letters are shown in **Supplementary Table S2**.

### Comparison of Discharge Letters

All letters underwent an initial independent screening by two study team members (J.T, S.R.G) for factual inaccuracy and grammatical errors. For letters generated by GPT-4, this included screening

for instances of hallucination by the language model. For cases 3 and 4, letters were also assessed on their attention to secondary problems beyond the prompted surgical issue by the presence or absence of follow-up recommendations for other medical issues found during the course of the patient's treatment.

A panel of five senior primary care physicians were presented with the discharge letters written by GPT-4 and junior clinicians in a blinded fashion. The participating physicians on the expert panel had an average clinical practice duration of 34.4 years. They compared the letters produced by GPT-4 and junior clinicians using a standardized rubric. Letters were evaluated on a 5-point Likert scale for completeness of information, conciseness, clarity, collegiality, whether follow-up care plans were articulated, and a single overall satisfaction question (**Figure 1**). While there are no validated tools for assessment of outpatient discharge letter quality, questionnaire items were selected based on elements endorsed by the TOCCC for inpatient summaries.

### *Data Analysis*

IBM SPSS Statistics version 26.0 (IBM Corp) was used for statistical analysis of quantitative data. Total scores and arithmetic means with standard deviation were calculated for all responses as a whole, and for responses to each individual item. As the data was non-normal in distribution (Shapiro-Wilk  $p < 0.05$ ), differences in mean were compared using the Mann-Whitney U test.

Figure 1: Scoring questionnaire

Question	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
1. Information about the patient's diagnosis and clinical course is satisfactory.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. The discharge summary is appropriately concise.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Information in the discharge summary is presented with clarity.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. The discharge summary is collegial.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. Recommendations for follow-up are clear.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. Overall, I would be happy to receive this discharge summary as the patient's next managing physician.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

## Results

Discharged letters generated by GPT-4 averaged 276 words (1462 characters), in comparison to 155 words (737 characters) for letters written by junior clinicians. No instances of hallucination and no grammatical errors were present in the letters generated by GPT-4. While no factual inaccuracies were found in the letters written by the junior clinicians, they made an average of 1.7 grammatical errors per letter.

Each panel member scored all 15 letters, for a total of 75 responses. Mean total score per letter was 22.99 (SD 6.26). GPT-4 scored higher on question items 1 (Information), 3, (Clarity), 4 (Collegiality), 5 (Follow-up plans), 6 (Overall satisfaction), and received higher total mean scores as compared to human subjects, but this difference was only statistically significant for Item 1. GPT-4 received lower scores on question item 2 (Conciseness). Results are shown in **Table 1**.

Item	GPT-4 (N=25)	Junior clinicians (N=50)	<i>p</i>
Information	4.32 (0.95)	3.70 (1.27)	0.03
Conciseness	3.60 (1.12)	3.64 (1.27)	0.71
Clarity	4.16 (0.94)	3.68 (1.24)	0.12
Collegiality	4.36 (1.00)	3.84 (1.22)	0.05
Recommendations	4.16 (1.03)	3.72 (1.13)	0.08
Overall satisfaction	3.96 (1.14)	3.62 (1.34)	0.36
Total	24.56 (5.13)	22.20 (6.66)	0.20

GPT-4 also had some unanticipated strengths when compared against junior clinicians. In particular, it demonstrated attention to secondary problems that were not explicitly included in the original prompt. In cases 3 and 4, GPT-4 provided follow-up recommendations based on the given fictional records for the incidental pancreatic lesion (“we recommend annual imaging studies to monitor the renal stone and the pancreatic lesion, in collaboration with the General Surgery team.”) and peri-operative hypertension (“he is recommended to continue with Amlodipine 5mg daily for his hypertension”) respectively. In contrast, both junior clinicians made mention of the pancreatic lesion and subsequent General Surgery referral, but did not provide follow-up recommendations to the primary care provider.

## Discussion

To our knowledge, this study represents the first blinded comparison of discharge letters written by an advanced LLM versus human physicians.

Our analysis showed that GPT-4 performed comparably against junior clinicians from an inpatient specialist department, with equivalent scores in clarity, collegiality, and ability to articulate follow-up recommendations. It also demonstrated the ability to generate a holistic patient summary with attention to secondary medical issues, and made fewer grammatical errors than human comparators. The higher score in information provision highlights the capability of language models with strong embeddings to capture and present important details from patient EMRs. However, this appears to have come at the expense of a relative lack of conciseness, as indicated by a lower score in this domain. More specific meta-prompting with instructions for word limits could overcome this limitation. Nevertheless, the importance of an LLM's ability to rapidly digest and summarize a large free-text corpus cannot be overstated, and offers the potential for a more consistent information relay between healthcare providers.

These findings echo prior work in the field comparing electronic discharge summaries against traditional dictated summaries. While earlier digitization efforts were limited by technological capabilities of the era, advances in natural language processing and linguistic ability allow modern language models to write fluently and mimic human prose convincingly. While the potential of language models in this area of healthcare has been discussed<sup>2</sup>, these first results prove that LLMs do, in actual fact, have the capability to automate time-consuming tasks for healthcare professionals. This is a transformative opportunity, and could be a useful tool in bridging communications between primary care providers and inpatient specialists by ensuring quality and accuracy of discharge letters. Clinicians burdened with documentation in addition to clinical duties stand to benefit immensely from automation of these tasks, allowing more time for patient contact and clinical care. Likewise, patients could benefit from consistent, timely handovers and a more robust system of continuity of care, instead of the vagaries of the junior clinicians writing their discharge letters. With appropriate refinement of prompt instructions, we postulate that this transformative benefit can extend beyond automating documentation work of physicians to other healthcare professionals e.g. nurses and pharmacists.

The exclusive reliance on fictional patient data in this study, while necessary for ethical and patient data considerations, might not capture real-world complexities in patient EMRs. We attempted to overcome this limitation by having the fictional EMRs written by physicians who also regularly see patients in the Urology outpatient clinic on a day-to-day basis and enter clinical notes into actual patient EMRs. The small number of panel assessors, as well as the use of clinical cases from a single surgical speciality, may reduce the generalisability of these findings. In addition, despite the advantages of LLM tools, misinformation phenomena such as hallucination and concept drift remain areas of concern. We propose that such an AI tool should be used to augment, not replace, human-written discharge letters, and that physician oversight should still be required before discharge documents are handed to patients. Users of LLMs as productivity aids in clinical medicine must also be aware of prevailing privacy protection policies, and closed-access LLM implementations (e.g. on a private server) may be needed for regulatory compliance. While automating discharge letter writing may undermine the clinical reasoning process that physicians undertake<sup>13</sup>, we believe that LLMs implemented with "human in the loop" can eventually exert a long term training effect, reinforcing these cognitive skill sets<sup>14</sup>.

Future research using actual patient EMRs will be necessary to confirm these preliminary findings.

Studies may be expanded to care across different medical specialties, to ascertain the ability of language models to comprehend abbreviations and concepts from different disciplines. Fine-tuning a language model on medical information and patient EMRs, and refining meta-prompts, are avenues to further improve the quality of discharge letters generated by LLMs, and to balance conciseness with information density.

### **Conclusion**

The emergence of AI in healthcare promises a paradigm shift in the way clinical medicine is practiced. Our study provides an insight into AI's capacity to optimize the continuity of patient care. As evaluation of language models for clinical applications continues to advance, we propose benchmarking performance against human counterparts to determine feasibility and assess output quality.

## References

1. Thirunavukarasu AJ, Ting DSJ, Elangovan K, et al. Large language models in medicine. *Nat Med*. 2023;29:1930–1940.
2. Patel SB, Lam K. ChatGPT: the future of discharge summaries? *Lancet Digit Health*. 2023;5:e107–e108.
3. Horwitz LI, Jenq GY, Brewster UC, et al. Comprehensive quality of discharge summaries at an academic medical center. *J Hosp Med*. 2013;8:436–443.
4. Rao P, Andrei A, Fried A, et al. Assessing quality and efficiency of discharge summaries. *Am J Med Qual*. 2005;20:337–343.
5. Weetman K, Spencer R, Dale J, et al. What makes a “successful” or “unsuccessful” discharge letter? Hospital clinician and General Practitioner assessments of the quality of discharge letters. *BMC Health Serv Res*. 2021;21:349.
6. Cresswell A, Hart M, Suchanek O, et al. Mind the gap: Improving discharge communication between secondary and primary care. *BMJ Qual Improv Rep*.;4 . Epub ahead of print September 2, 2015. DOI: 10.1136/bmjquality.u207936.w3197.
7. Maslove DM, Leiter RE, Griesman J, et al. Electronic versus dictated hospital discharge summaries: a randomized controlled trial. *J Gen Intern Med*. 2009;24:995–1001.
8. Bernstein IA, Zhang YV, Govil D, et al. Comparison of Ophthalmologist and Large Language Model Chatbot Responses to Online Patient Eye Care Questions. *JAMA Netw Open*. 2023;6:e2330320.
9. Singhal K, Azizi S, Tu T, et al. Large language models encode clinical knowledge. *Nature*. 2023;620:172–180.
10. OpenAI. GPT-4 Technical Report. *arXiv [cs.CL]*. Available from: <http://arxiv.org/abs/2303.08774>. 2023.
11. Bender EM, Gebru T, McMillan-Major A, et al. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? . In: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. New York, NY, USA: Association for Computing Machinery; 2021:610–623.
12. Zhou C, He J, Ma X, et al. Prompt Consistency for Zero-Shot Task Generalization. *arXiv [cs.CL]*. Available from: <http://arxiv.org/abs/2205.00049>. 2022.
13. Preiksaitis C, Sinsky CA, Rose C. ChatGPT is not the solution to physicians’ documentation burden. *Nat Med*. 2023;29:1296–1297.
14. Sutton RT, Pincock D, Baumgart DC, et al. An overview of clinical decision support systems: benefits, risks, and strategies for success. *NPJ Digit Med*. 2020;3:17.

## Supplementary Material

## Supplementary Figure S1: Prompt instructions

You are a physician assistant in a Urology clinic. You will write a discharge summary of a patient's medical condition. This patient is being discharged from Specialist Urological care, and the discharge summary is meant to be read by the General Practitioner to whom the patient is being discharged.

Make use of all the documents provided, which include original referral letters, clinic consultation notes, and discharge summaries from related admissions. You will need to extract the patient's background information and past medical history, the relevant urological issue for which he is being treated, and the course of his condition. Do not include any information that is not already in the documents.

Desired format:

1. Patient biodata, background information, and a brief summary of significant past medical history. Itemise the patient's list of past medical history.
2. Summary of original referral and/or presenting complaint
3. List of issues managed by Urology over the duration of urological follow-up in the Urology outpatient clinics, with relevant investigations and treatment rendered summarised under each issue. Include dates where relevant.
4. Discharge information including new medications, medication changes, and recommended follow-up plans

To assist you to prepare this discharge summary, I will give you a series of documents, starting with the original referral letter, and followed by each visit's clinic entry or discharge summary, up until the point of discharge. Write in prose form. The letter is meant to be addressed to the original referring General Practitioner. The letter should be cordial, concise, and should not exceed half a page of text unless necessary. Start your letter with "Dear Colleague, thank you for your referral of this patient."

Supplementary Table S2: Sample fictional clinical records and discharge letters

Sample consultation notes from Case 1	GPT-4	Medical Officer 1	Medical Officer 2
<p>Uro Consult 8 June 2018</p> <p>65/M NKDA Non-smoker Nil PMH</p> <p>Referred for elevated PSA 5.3 Feb 2018. Also c/o voiding LUTS - poor stream, hesitancy. No frequency/urgency. No GH/dysuria No aggravating factors. No fhx of renal/prostate/bladder CA</p> <p>OE Abdo soft, non-tender Penis, scrotum, testes NAD DRE - benign, 3FB</p> <p>BSUS b/l no HN PV 45cc IPPG1</p> <p>Offered further evaluation of elevated PSA with MRI prostate - not keen, prefers to trend PSA. Understands risk of missed CA diagnosis.</p> <p>P: UFEME, Urine c/s Alfuzosin - counseled re: giddiness TCU 3/12 UFRU PSA OA Uro Consult 5 Jan 2020</p> <p>MRI Prostate Dec 2019: PIRADS 3 x1 0.4cm Left Anterior PZ, Prostate volume 75cc.</p> <p>Offered biobot prostate biopsy - keen. Counselling re: risks of bleeding, infection, ARU.</p> <p>P: List for biobot prostate biopsy on 17 Feb 2020 as DS under GA PAAC Urine c/s PP Uro Consult 8 Mar 2023 Well Happy w LUTS</p> <p>UF 12/450 RU 13ml</p> <p>PSA 1.25 (x2=2.5)</p> <p>Patient request for transfer of care to OPS as he feels his condition is stable.</p>	<p>Dear Colleague,</p> <p>Thank you for your referral of Mr. Wong Chin Jin, a 65-year-old gentleman, who initially presented to our Urology service with elevated Prostate-Specific Antigen (PSA) levels and lower urinary tract symptoms.</p> <p>Over the course of five years, we closely monitored Mr. Wong's PSA levels and urinary symptoms. Initially, he was hesitant about further evaluation of his elevated PSA levels, opting instead for symptomatic management and PSA trending. We instituted Alfuzosin, and he reported satisfaction with the management of his symptoms.</p> <p>In 2019, Mr. Wong consented to an MRI of the prostate. The results showed a PIRADS 3 x1 0.4cm lesion in the Left Anterior Peripheral Zone. Consequently, he underwent a Transperineal Biobot Prostate Biopsy in February 2020, which unfortunately led to a post-operative complication of Acute Retention of Urine (ARU). However, he recovered well, and the histopathological examination confirmed the presence of Benign Prostatic Hyperplasia (BPH).</p> <p>From March 2020, Mr. Wong was started on Finasteride, in addition to the ongoing Alfuzosin, to manage his BPH. Subsequent follow-ups demonstrated satisfaction with his symptom control, and his PSA levels notably decreased to around 2.5.</p> <p>As of his last consultation in March 2023, Mr. Wong feels his condition is stable and has requested to be transferred to your care in the Outpatient Department. We recommend that he continues his current medications of Alfuzosin and Finasteride for BPH. We have left an open date for him in Urology, should he require any specialized care in the future.</p> <p>Thank you for entrusting us with Mr. Wong's urological care.</p> <p>Sincerely, [Your Name]</p>	<p>Dear Senior Colleague,</p> <p>Thank you for your referral for Mr Wong who is known to Uro for BPH stable on alfuzosin and finasteride, currently keen for f/u in your clinic setting. He was initially referred for Raised PSA of 5.3 with the following management.</p> <p>65/M NKDA Non-smoker</p> <p>Nil PMH</p> <p>Uro Issues 1. Raised PSA - PSA: 5.3 (Feb 2018) &gt; 5.5 (Sept 2018) &gt; 5.9 (Sept 2019) &gt; PSA 1.30x2=2.6 (4 Mar 2022) &gt; PSA 1.25x2=2.5 (8 Mar 2023) - Dec 2019: MRI Prostate - PIRADS 3 x1 0.4cm Left Anterior PZ, Prostate volume 75cc. - 17 Feb 2020 - Biobot Prostate Biopsy, Histo: BPH</p> <p>2. LUTS 2' to BPH - Dec 2019: MRI Prostate - PIRADS 3 x1 0.4cm Left Anterior PZ, Prostate volume 75cc - 8 June 2018 - Started on alfuzosin - 17 Mar 2020 - started on finasteride - 8 March 2023: UF 12/450 RU 13ml</p> <p>We would appreciate you continued care of this patient and to prescribe his medications as required. He has been given an open date from Urology. Thank you for your continued care of this patient.</p> <p>Regards</p>	<p>Dear Colleague, thank you for your referral of this patient with elevated PSA 5.3 done as part of health screening and was diagnosed to have BPH. He was initially started on alfuzosin for his associated voiding LUTs and declined MRI prostate for further evaluation from June 2018 to Sep 2019. His PSA trend during the corresponding period was 5.3 &gt; 5.5 &gt; 5.9.</p> <p>He was eventually agreeable for MRI prostate which showed PIRADS 3 x1 0.4cm Left Anterior PZ, prostate volume 75cc. Transperineal biobot biopsy was eventually done on 17/2/20 which showed BPH. Finasteride was added on top of alfuzosin in March 2020 with improvement in his LUTs. His condition has since has been stable from 2020 to 2023 with the following PSA trend: 2.5 &gt; 2.6 &gt; 2.5. He eventually requested for transfer of care to OPS in view of his stable condition.</p> <p>Kindly assist in following up for his BPH and continue his alfuzosin and finasteride.</p> <p>Thank you.</p>



<p>P: Open date Uro Cont. alfuzosin and finasteride Memo to OPS to continue BPH medications</p> <p><b>Sample consultation notes from Case 1</b></p> <p>Uro Consult 28 Dec 2022</p> <p>32/M NKDA Non-smoker Works office desk job (IT) Nil PMH</p> <p>R ureteric colic Seen A&amp;E 2 weeks ago Still having pain intermittently Loin to groin, colicky One episode gross hematuria which resolved No dysuria No fever No prior stone hx No fmhx cancer/stones</p> <p>A&amp;E UFEME 40/10/0 KUB XR ?right ureteric stone</p> <p>OE Abdo soft, non-tender Renal punch neg Penis, scrotum, testes NAD DRE - benign, 2.5FB</p> <p>P: UFEME, urine culture CT IVP TCU 2/52 after Discharge Summary</p> <p>Patient Name: Aaron Lee Admission Date: 6 March 2023 Discharge Date: 6 March 2023 Admission Type: Day Surgery</p> <p>Elective admission for Right Ureteroscopy and laser lithotripsy.</p> <p>Patient underwent Right Ureteroscopy and laser lithotripsy on 6 Mar 2023. Peri-operatively was noted to have hypertension and started on amlodipine 5mg OM.</p> <p>Discharged well and stable on 6 Mar 2023.</p> <p>List for DJ stent removal on 18 Mar 2023 in Urology clinic. TCU OPS in 1/12 to review hypertension</p> <p>Discharge medications: Amlodipine 5mg OM 1 month</p>	<p><b>GPT-4</b></p> <p>Dear Colleague,</p> <p>Thank you for your referral of Mr. Aaron Lee, a 32-year-old IT professional with no prior medical history. He was initially seen in A&amp;E for right-sided ureteric colic on 17 Dec 2022.</p> <p>Mr. Lee presented with intermittent loin to groin pain, with one episode of gross hematuria, which later resolved. A subsequent CT IVP revealed a 3mm right distal ureteric stone. Given the persistence of the stone on follow-up imaging, despite an initial attempt at conservative management with Medical Expulsive Therapy (MET), he underwent a successful Right Ureteroscopy and laser lithotripsy on 6 Mar 2023.</p> <p>During his day surgery admission for the procedure, he was noted to have hypertension and was started on Amlodipine 5mg once daily. A follow-up double J (DJ) stent removal was conducted uneventfully on 18 Mar 2023. A follow-up XR KUB on 28 June 2023 revealed no residual stones.</p> <p>Upon discharge, he is recommended to continue with Amlodipine 5mg daily for his hypertension. He had failed to attend his scheduled follow-up at the Outpatient Services (OPS) for his hypertension in March. Thus, I would appreciate if this could be followed up in your care.</p> <p>Please also advise him to return to Urology services if he experiences similar symptoms in the future. Thank you again for your referral.</p> <p>Kind regards,</p> <p>[Your Name] [Your Title]</p>	<p><b>Medical Officer 1</b></p> <p>Dear Senior Colleague,</p> <p>Thank you for your referral for Mr Lee who is known to Uro for Right Ureteric stone s/ p Right Ureteroscopy and laser lithotripsy, currently keen for f/u in your clinic setting. He was also noted to have newly diagnosed HTN started on amlodipine 5mg OM. He was initially referred for Right Ureteric Colic with the following management.</p> <p>32/M NKDA Non-smoker Works office desk job (IT)</p> <p>PMH 1. HTN - started on Amlodipine 5mg OM</p> <p>Uro Issues 1. Right ureteric stone s/p Right Ureteroscopy and laser lithotripsy on 6 March 2023 - Initial presentation of Right ureteric colic on 28 Dec 2022 - DRE - benign, 2.5FB - 15 Jan 2023: CT IVP: 3mm right distal ureteric stone - s/p MET (Tamsulosin) - 28 Feb 2023: CT KUB: 3mm right distal ureteric stone, no migration of stone. Mild right hydronephrosis seen - s/p Right Ureteroscopy and laser lithotripsy on 6 March 2023 and flex cysto and stent removal on 18 March 2023 - 28 June 2023: XR KUB - no stone seen</p> <p>Other issues 1.HTN - started on Amlodipine 5mg OM on 6 Mar 2023</p> <p>We would appreciate you continued care of this patient and to prescribe his medications as required. He has been given an open date from Urology. Thank you for your continued care of this patient.</p> <p>Regards</p>	<p><b>Medical Officer 2</b></p> <p>Dear Colleague, thank you for your referral of this patient with ureteric colic. The following investigations were done: Urine cultures negative CT IVP: 3mm right distal ureteric stone</p> <p>He was initially keen for Medical Expulsive Therapy and was given a 6 week course of tamsulosin from Jan 2023 to Feb 2023. A repeat CT KUB showed persistence of 3mm right distal uretic stone with no migration and newly developed right hydronephrosis. He was eventually keen for Right ureteroscopy and laser lithotripsy which was done on 6 March 2023 with removal of DJ stent (inserted intraoperatively) on 18 March 2023. A repeat XR KUB in June 2023 did not show presence of any stones.</p> <p>He was also noted to have hypertension on the day of uretoscopy and was started on amlodipine 5mg OM. However he defaulted follow up with OPS for his hypertension.</p> <p>Kindly assist in following up with his hypertension. Return advice has been given to patient.</p> <p>Thank you!</p>
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

<p>Augmentin 1g BD 5 days Paracetamol 1g QDS PRN 7 days Uro Procedure 18 Mar 2023</p> <p>Flex scope and R DJ stent removal performed. Stent removed in entirety - tip intact.</p> <p>P: TCU 3/12 XR KUB OA Uro Consult 28 June 2023</p> <p>Well XR KUB - no stone seen</p> <p>Did not go to OPS for Hypertension follow up in March.</p> <p>P: Discharge to OPS - memo to followup HTN Open date Uro Return advice given</p>			
------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--	--	--