Generating a Benzodiazepine Taper Using a Large Language Model: Feasibility Study

Obinna Ekekezie

Submitted to: JMIR Formative Research
on: August 19, 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 it's 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 expressively prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Generating a Benzodiazepine Taper Using a Large Language Model: Feasibility Study

Obinna Ekekezie[1, 2] MD, AB

[1]Cambridge Health Alliance Department of Psychiatry Cambridge US
[2]Harvard Medical School Boston US

Corresponding Author:
Obinna Ekekezie MD, AB
Cambridge Health Alliance
Department of Psychiatry
1493 Cambridge St.
Cambridge
US

Abstract

Exploring the feasibility of using artificial intelligence (AI), specifically large language models (LLMs), to automate the generation of benzodiazepine tapering plans.

(JMIR Preprints 19/08/2024:65570)
DOI: https://doi.org/10.2196/preprints.65570

Preprint Settings

Original Manuscript

Generating a Benzodiazepine Taper Using a Large Language Model: Feasibility Study

Obinna Ikechukwu Ekekezie, M.D.[a,b]

[a] Cambridge Health Alliance, Cambridge, MA, United States of America
[b] Harvard Medical School, Boston, MA, United States of America
Corresponding author's contact information: oekekezie@challiance.org

Manuscript word count: 621

Abstract: Exploring the feasibility of using artificial intelligence (AI), specifically large language models (LLMs), to automate the generation of benzodiazepine tapering plans.

Introduction

According to the 2022 National Survey on Drug Use and Health survey conducted by the Substance Abuse and Mental Health Services Administration, about 1.4% of American adults (>18 years old) endorsed having misused benzodiazepines in the past year [1]. For those who have been taking benzodiazepines daily for a month or more, it is important to taper off the medication and avoid abrupt discontinuation as it could precipitate a severe or even life threatening withdrawal [2]. Primary care providers and psychiatric prescribers often find it challenging to deprescribe benzodiazepines in these situations citing insufficient time (in terms of visit duration and lack of appointment availability) and lack of access to specialists or consultants with more experience in tapering patients off of sedatives [3].

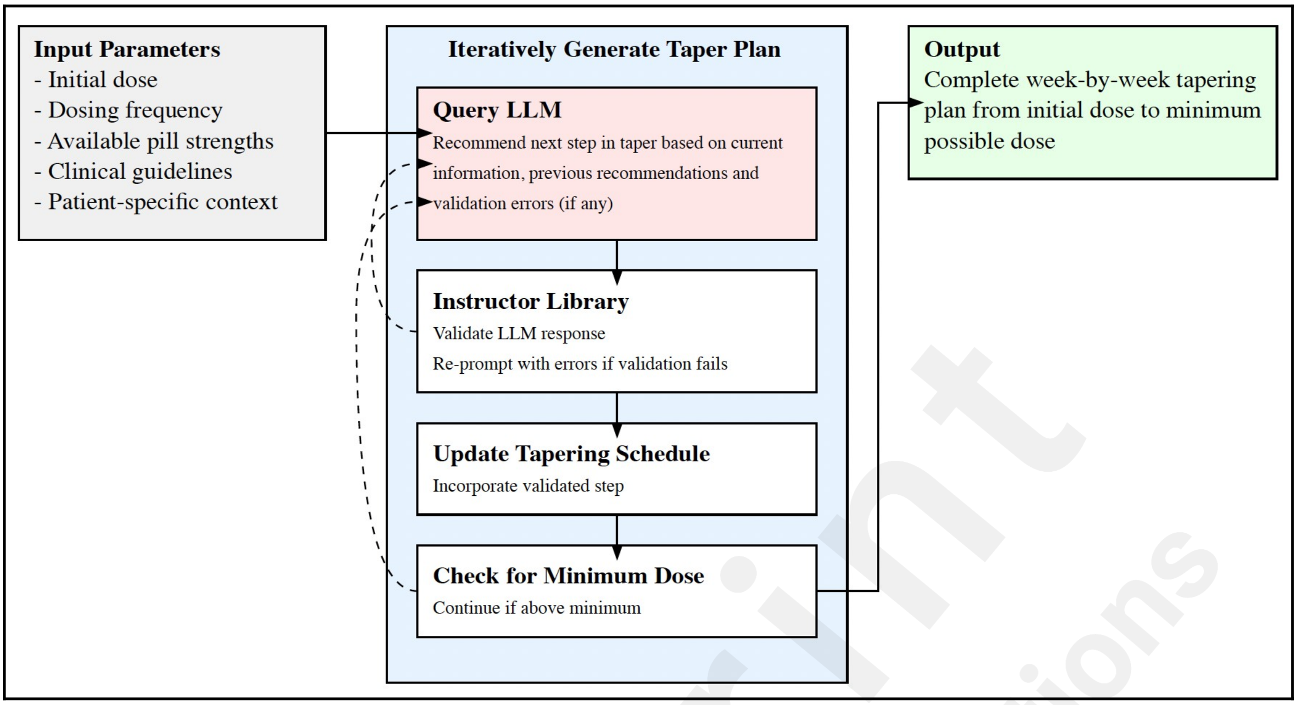
Previous research has shown that clinical decision support (CDS) tools can save clinicians time [4]. In addition, recent feasibility research has shown that, when grounded in established clinical guidelines, large language models (LLMs) may have potential in augmenting clinical decision making when prescribing psychiatric medications [5]. This research project sought to explore the feasibility of using an LLM to generate a week-by-week benzodiazepine tapering plan.

Methods

To avoid using sensitive patient information, a synthetic dataset of 20 different clinical scenarios was generated using an LLM: 10 examples were used to develop the algorithm for iteratively generating tapering plans *(dev)* while the other 10 were reserved for assessing the algorithm’s generalizability *(test)*.

The algorithm, written in Python, iteratively queries an LLM, Anthropic’s Claude 3.5 Sonnet (“claude-3-5-sonnet-20240620”), to recommend the next step in a benzodiazepine taper until the minimum possible daily dose is reached resulting in the creation of a complete week-by-week tapering plan (Figure 1). The LLM considers the initial dose, dosing frequency, available pill strengths, clinical guidelines, and patient-specific context. The Instructor Python library was used to validate the LLM's responses and prompt for corrections when necessary (see eMethods in Multimedia Appendix).

Figure 1. Overview of Tapering Plan Algorithm


Input Parameters
- Initial dose
- Dosing frequency
- Available pill strengths
- Clinical guidelines
- Patient-specific context
Iteratively Generate Taper Plan
Query LLM
Recommend next step in taper based on current information, previous recommendations and validation errors (if any)
Instructor Library
Validate LLM response
Re-prompt with errors if validation fails
Update Tapering Schedule
Incorporate validated step
Check for Minimum Dose
Continue if above minimum
Output
Complete week-by-week tapering plan from initial dose to minimum possible dose

Performance metrics included the mean number of attempts required per step, percentage of steps requiring multiple attempts, maximum attempts per step, and costs per generated taper plan. The generated tapers were assessed by an LLM-as-a-judge evaluating whether predefined criteria were met: correctly identifying weekly dose reduction targets, justifying exceeding target reductions, adhering to dosing frequency guidelines, justifying frequency changes, and correctly allocating the largest dose to the end of the day. OpenAI's GPT-4o ("gpt-4o-2024-05-13") was used as the judge to evaluate Claude's performance to reduce bias (see eMethods in Multimedia Appendix).

Results

The LLM required multiple attempts to generate the next step about 25% of the time, with some steps needing up to 4 attempts (Table 1). On average, generating each taper cost about $0.50 to $0.60 based on Claude's current pricing.

Table 1. Evaluation of Tapering Plan Algorithm

Evaluation Criterion	Description	Dev	Test
Attempts required per step, mean (SD)	Attempts required to produce a single valid step in the taper plan	1.28 (0.53) N=123	1.26 (0.48) N=119
Multi-attempt percentage	Percentage of steps requiring multiple attempts to generate output that passed validation	25% (31/123)	24% (29/119)
Maximum attempts	Highest number of attempts required to produce a single valid step in the taper plan	4	3

Cost per generated taper plan, mean (SD), $USD	Cost to generate a complete week-by-week taper plan	$0.57 ($0.43)	$0.49 ($0.28)
Assessed by the LLM-as-a-Judge			
Correctly identifies weekly dose reduction target	Identifies the correct guideline-recommended target weekly dose reductions. The rationale for each week should identify the correct target based on the clinical context. Each assertion about dose reduction percentages must be factually and mathematically correct.	41% (50/123)	41% (49/119)
Correctly justifies exceeding target reductions	When exceeding the guideline-recommended target, the rationale explicitly acknowledges this and provides a valid justification based on at least one of: 1) reaching the minimum possible dose, 2) practical necessity due to available doses, and/or 3) clinical context.	15% (11/72)	40% (28/70)
Adherence to frequency guidelines	The dosing frequency either remains the same or decreases from one week to the next. It never increases and doesn't change from once daily (bedtime) to once daily (daytime).	98% (120/123)	99% (118/119)
Justified frequency changes	When changing dosing frequency, the rationale explicitly acknowledges this and provides a valid justification based on at least one of: 1) reaching minimum possible dose, 2) practical necessity due to available doses, 3) clinical context, and/or 4) pharmacokinetics.	86% (19/22)	83% (15/18)
Correct end-of-day dose allocation	The plan allocates the largest dose to the end of the day whenever there is an unequal allocation of doses.	94% (72/77)	100% (49/49)

The LLM correctly applied guideline-based target weekly dose reductions only about 40% of the time. When recommending reductions that did not align with guidelines, it often failed to acknowledge or justify the deviation.

The LLM largely respected guidelines for maintaining or decreasing dosing frequency. When changing frequency, it provided reasonable justification in most instances. The LLM successfully allocated larger doses towards the end of the day in the majority of cases in which the total daily dose had to be unequally divided throughout the day.

Discussion

This feasibility study suggests that, with refinements, it may be possible to leverage an LLM to generate week-by-week benzodiazepine tapering plans cost-effectively. However, the study is limited by using another LLM to evaluate the generated plans' quality, rather than human subject matter experts.

Future research should involve human expert evaluation of the LLM-generated plans. Additionally, considering that clinicians may prefer to draft tapers for a few weeks at a time so they can adjust them according to patients' withdrawal symptoms, exploring ways to modify the LLM-generated plans at any given step could be valuable next steps in this area of research.

Acknowledgements

Ikechukwu Ekekezie, MD and Hsiang Huang, MD, MPH for providing constructive feedback regarding the early conceptualization of this feasibility research study.

References

1. Substance Abuse and Mental Health Services Administration. 2022 National Survey on Drug Use and Health. Accessed July 27, 2024. https://www.samhsa.gov/data/release/2022-national-survey-drug-use-and-health-nsduh-releases

2. Ogbonna CI, Lembke A. Tapering Patients Off of Benzodiazepines. *Am Fam Physician*. 2017;96(9):606-610.

3. Hawkins EJ, Lott AM, Danner AN, et al. Primary Care and Mental Health Prescribers, Key Clinical Leaders, and Clinical Pharmacist Specialists' Perspectives on Opioids and Benzodiazepines. *Pain Medicine*. 2021;22(7):1559-1569. doi:10.1093/pm/pnaa435

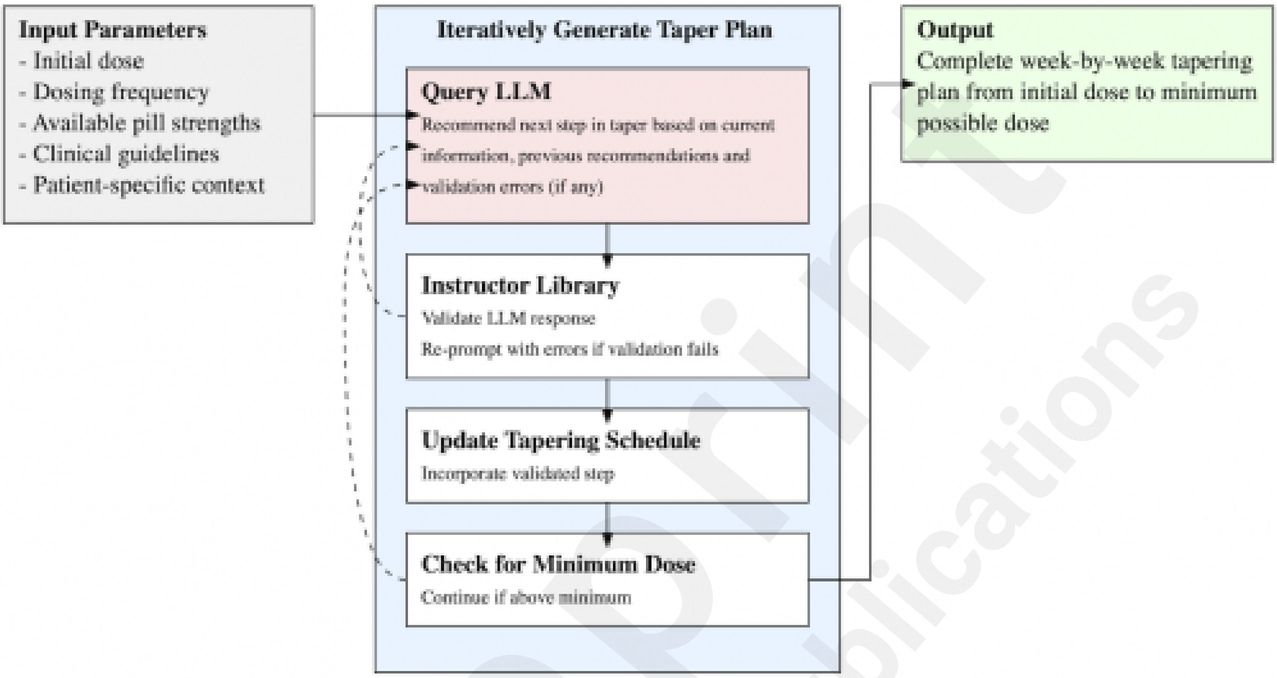
4. Wagholikar KB, Hankey RA, Decker LK, et al. Evaluation of the Effect of Decision Support on the Efficiency of Primary Care Providers in the Outpatient Practice. *J Prim Care Community Health*. 2015;6(1):54-60. doi:10.1177/2150131914546325

5. Perlis RH, Goldberg JF, Ostacher MJ, Schneck CD. Clinical decision support for bipolar depression using large language models. *Neuropsychopharmacol*. 2024;49(9):1412-1416. doi:10.1038/s41386-024-01841-2

Supplementary Files

Figures

An overview of the generative tapering plan algorithm.


Input Parameters
- Initial dose
- Dosing frequency
- Available pill strengths
- Clinical guidelines
- Patient-specific context
Iteratively Generate Taper Plan
Query LLM
Recommend next step in taper based on current information, previous recommendations and validation errors (if any)
Instructor Library
Validate LLM response
Re-prompt with errors if validation fails
Update Tapering Schedule
Incorporate validated step
Check for Minimum Dose
Continue if above minimum
Output
Complete week-by-week tapering plan from initial dose to minimum possible dose

Multimedia Appendixes