

RESEARCH

Open Access



Challenging cases of hyponatremia incorrectly interpreted by ChatGPT

Kenrick Berend^{1*}, Ashley Duits^{2,3} and Reinold O. B. Gans^{1,4}

Abstract

Background In clinical medicine, the assessment of hyponatremia is frequently required but also known as a source of major diagnostic errors, substantial mismanagement, and iatrogenic morbidity. Because artificial intelligence techniques are efficient in analyzing complex problems, their use may possibly overcome current assessment limitations. There is no literature concerning Chat Generative Pre-trained Transformer (ChatGPT-3.5) use for evaluating difficult hyponatremia cases. Because of the interesting pathophysiology, hyponatremia cases are often used in medical education for students to evaluate patients with students increasingly using artificial intelligence as a diagnostic tool. To evaluate this possibility, four challenging hyponatremia cases published previously, were presented to the free ChatGPT-3.5 for diagnosis and treatment suggestions.

Methods We used four challenging hyponatremia cases, that were evaluated by 46 physicians in Canada, the Netherlands, South-Africa, Taiwan, and USA, and published previously. These four cases were presented two times in the free ChatGPT, version 3.5 in December 2023 as well as in September 2024 with the request to recommend diagnosis and therapy. Responses by ChatGPT were compared with those of the clinicians.

Results Case 1 and 3 have a single cause of hyponatremia. Case 2 and 4 have two contributing hyponatremia features. Neither ChatGPT, in 2023, nor the previously published assessment by 46 clinicians, whose assessment was described in the original publication, recognized the most crucial cause of hyponatremia with major therapeutic consequences in all four cases. In 2024 ChatGPT properly diagnosed and suggested adequate management in one case. Concurrent Addison's disease was correctly recognized in case 1 by ChatGPT in 2023 and 2024, whereas 81% of the clinicians missed this diagnosis. No proper therapeutic recommendations were given by ChatGPT in 2023 in any of the four cases, but in one case adequate advice was given by ChatGPT in 2024. The 46 clinicians recommended inadequate therapy in 65%, 57%, 2%, and 76%, respectively in case 1 to 4.

Conclusion Our study currently does not support the use of the free version ChatGPT 3.5 in difficult hyponatremia cases, but a small improvement was observed after ten months with the same ChatGPT 3.5 version. Patients, health professionals, medical educators and students should be aware of the shortcomings of diagnosis and therapy suggestions by ChatGPT.

*Correspondence:
Kenrick Berend
kenber2@me.com

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Key learning points

Proper evaluation of hyponatremia cases is difficult for physicians, but currently also for ChatGPT 3.5.

Keywords Ecstasy, Hyponatremia, Low osmol intake, Mineralocorticoid deficiency, SIADH

Introduction

Assessing hyponatremia in patients is one of the most complex challenges in medicine [1, 2, 3, 4, 5, 6, 7]. For more than half a century we rely on an imperfect approach, centralized on the volume status in the non-edematous patient with hyponatremia. One of the problems is that it is often difficult to differentiate a volume depleted hyponatremic state from a dilutional hyponatremic condition [1, 4, 5, 6, 7, 8, 9, 10, 11]. Also, a single cause of hyponatremia may be present in less than 50% of patients presenting in the emergency department, limiting the use of current hyponatremia evaluation flow charts [1, 7, 8, 9, 10].

Hyponatremia affects up to 30% of hospitalized patients and errors in diagnosis may lead to adverse outcome [11]. Nevertheless, its evaluation remains a significant medical problem. With the explosive development of deep learning, artificial intelligence is rapidly becoming a potential valuable assistant in many applications in medicine, such as diagnosing difficult medical cases. Furthermore, patients with hyponatremia are often used for students' evaluation assignment during clerkships. In addressing diagnostic and treatment recommendations artificial intelligence is increasingly used by students and physicians [12, 13, 14, 15, 16, 17, 18, 19]. Large language models have shown to be useful in multiple-choice and open-ended medical reasoning examinations and are gradually integrated into many health care settings. It remains unknown whether platforms like Chat Generative Pre-Trained Transformer (ChatGPT) may be particularly helpful in complex hyponatremia cases to improve diagnosis and therapy. Studies using machine learning showed it is possible to predict the onset and treatment response of hyponatremia in certain hospitalized patients [19], but there have not been any published reports of its evaluation in diagnosing the cause of hyponatremia by the free version of ChatGPT-3.5. To evaluate this option, the same four complex hyponatremia cases were entered in the free version of ChatGPT version 3.5 in 2023 and 2024 [12].

Methods

Four case vignettes used were adapted from a study by Hoorn et al. [1]. All 4 cases were originally evaluated by 46 physicians (27 residents, 6 fellows, 13 staff physicians in internal medicine specialties) in Canada, the Netherlands, South-Africa, Taiwan, and USA. The physicians had been provided with current clinical diagnostic

algorithms (CDA) for hyponatremia to suggest definitive diagnoses and management. Only 10% of all physicians reached a correct diagnosis using these CDA's.

With the request to suggest diagnosis and therapy, all cases [1] were entered into ChatGPT-3.5 (powered by OpenAI via Platform for Open Exploration) [12] on December 28, 2023 and September 3, 2024. Each case was run in a separate ChatGPT chat to prevent the model applying any "learning" to subsequent cases. The 2023 and 2024 ChatGPT responses were copied, analyzed and compared with the comment from the original analysis of each case [1].

Results

The answers obtained by ChatGPT were compared with those by the physicians.

Case 1 [1]

A 73-year-old woman with nausea and vomiting for 3 weeks had lost 3.5 kg in weight over the past several months. She was not taking any medication. Physical examination noted a normal pulse rate and blood pressure without orthostatic changes. On admission plasma sodium was 123 mmol/l, potassium 4.5 mmol/l, glucose 4.3 mmol/l, creatinine 83 μ mol/l and thyroid stimulating hormone was 2.9 mIU/l. A spot urine sodium was 75 mmol/l [1, 20].

Hoorn et al.: [1] The diagnostic problem was Addison's disease in absence of the alleged hallmark of low extracellular fluid volume and hyperkalemia. Nausea and vomiting were caused by Addison's disease. When adrenal glands produce insufficient cortisol and aldosterone it is called Addison's disease, or primary adrenal insufficiency [21]. The secondary form develops when the pituitary gland causes the cortisol deficiency due to inadequate adrenocorticotrophic hormone (ACTH) production by a disease of the pituitary gland itself. Tertiary adrenal insufficiency is related to inadequate ACTH production by the hypothalamus as the cause of the low cortisol level, most often due to exogenous steroids. Primary adrenal insufficiency, unlike the other forms, often presents with dehydration, hyperkalemia, and/or skin hyperpigmentation [5, 7, 9].

In case 1, the attending physicians made a wrongful diagnosis of a syndrome of inappropriate ADH secretion (SIADH) [1, 20]. Fluid restriction obviously could not improve the mineralocorticoid deficient hyponatremia, and the patient died 2 weeks after admission. Post-mortem evaluation revealed adrenal failure. Addison's disease

Table Summary responses ChatGPT and clinicians in 4 patients with complex hyponatremia cases (adapted from Hoorn et al.) [1]

| case | Diagnostic and therapeutic aspects of the case | Did ChatGPT recognize the diagnosis and/or recommend dangerous therapy in 2023 and/or 2024? | Did clinicians recognize the diagnosis and/or recommend dangerous therapy? | Comment |
|------|---|---|--|--|
| 1 | Normokalaemia Addison's disease with apparent normovolaemia Treatment: saline and steroids. | Diagnosis missed in 2023 and 2024. No therapy recommended in 2023. In 2024 inappropriate treatment with fluid restriction suggested due to diagnostic error (SIADH). | Addison's disease recognized by 11%. Misdiagnosed as SIADH by 48% Inappropriate water restriction recommended by 65% | ChatGPT made no mentioning of the hyponatremia in 2023, but dangerous therapy suggestion because of wrong diagnosis (SIADH) was made in 2024. Diagnosis missed by 48% of clinicians. 65% of the clinicians advised a dangerous treatment with fluid restriction. |
| 2 | First and main problem: low osmol intake with a high risk of developing osmotic demyelination syndrome (ODS) with saline infusion. Second and concurrent diagnosis: Addison's disease Treatment: half-isotonic saline and plasma sodium should increase only 4 to 6 mmol/l in the first 24 h. | The main diagnostic aspect of low osmol intake was not recognized in 2023 or 2024. Addison's disease diagnosed in 2023 and 2024. Inappropriate treatment with normal saline recommended in 2023 and 2024. | Main diagnostic aspect of low osmol intake was not recognized. Addison's disease diagnosed by 19% of clinicians. Inappropriate treatment recommended by 57% of clinicians | Diagnosis of chronic hyponatremia with a presumed low osmol intake increases the risk of ODS not recognized by clinicians and ChatGPT in 2023 and 2024. Addison's disease missed by 81% of clinicians but diagnosed by ChatGPT in 2023 and 2024. Dangerous treatment with isotonic saline recommended by ChatGPT and most of the clinicians. |
| 3 | Hyponatremia was the result of low protein and salt intake with moderate water intake. If the patient is asymptomatic, fluid restriction should be initiated. Inappropriate treatment with isotonic saline causes overcorrection of sodium. | Diagnosis not recognized in 2023. In 2024 initial appropriate discussion of vegan lifestyle, with low-salt diet and high-water intake. However, in the same discussion diagnosis switched to primary polydipsia or exercise-associated hyponatremia. Dangerous treatment with normal saline recommended in 2023, but proper advice of fluid restriction was made in 2024. | Diagnosis not recognized by any of the clinicians. Wrongly diagnosed as psychogenic polydipsia by (53%) and SIADH by (12%) Appropriate fluid restriction was recommended by 80%. Dangerous treatment with normal saline recommended by 2% | Diagnosis not recognized by clinicians or ChatGPT (2023). In 2024 ChatGPT proposed the correct diagnosis and treatment. High risk treatment with isotonic saline recommended by ChatGPT in 2023 but proper treatment in 2024. Most of the clinicians recommended appropriate fluid restriction for other reasons, because of errors in diagnosis. |
| 4 | First and main problem. Temporarily increased plasma sodium due to seizure masking the severity of hyponatremia. Clue: discrepancy between the measured plasma sodium and the severity of the symptoms. Second problem. Severe hyponatremia (masked by seizure) caused by ecstasy and water intake. Treatment: because of acute symptomatic hyponatraemia, NaCl 3% is recommended. | The role of seizure on hyponatremia was not recognized in 2023 or 2024, but in both years the role of ecstasy was recognized. Inappropriate fluid restriction was recommended in 2023 and 2024. | No, seizure due to "masked" hyponatremia not recognized. Wrongly diagnosed as SIADH (11%), water ingestion (13%), and sodium loss in sweat (28%). Proper administration with hypertonic saline recommended by 24%. | The point in this case is that seizures can raise the plasma sodium acutely by 10–15 mmol/l was not recognized either by ChatGPT or clinicians [8]. Unsafe management suggestion by ChatGPT in 2023 and 2024 and inadequate management by most clinicians. |

was not suspected by the attending physicians because misleading similarities with SIADH such as absent clinical signs of a contracted extracellular volume (ECF), normokalaemia and a urine sodium of 75 mmol/l [20]. Adrenal insufficiency is one of the most important disorders to exclude before diagnosing SIADH, as fluid restriction and failure to initiate hydrocortisone treatment is hazardous [1, 5, 18, 19, 20, 21, 22]. Because there is no simple laboratory test to confirm SIADH, one should always evaluate the known SIADH criteria such as, low serum uric acid, normal serum cortisol, normal acid-base

parameters and the absence of other causes of hyponatremia [1, 4, 7, 23, 24].

The 46 physicians. *Addison's disease was missed by 89% of clinicians and hazardous water restriction was recommended by 65% (table) [1].*

ChatGPT (2023): *Given the patient's age, the chronic nature of her symptoms, and the weight loss, it's important to consider gastrointestinal, renal, or malignancy-related causes.*

ChatGPT (2024): *Given the clinical picture, this appears to be a case of euvolemic hyponatremia. The absence of*

medications and her normal physical examination suggest that common drug-induced causes (like diuretics) and fluid overload states (e.g., heart failure, cirrhosis) are unlikely. Her urine sodium is elevated at 75 mmol/L, which suggests SIADH. Normal potassium, glucose, creatinine, and TSH, ruling out adrenal insufficiency, hypothyroidism, and kidney issues as causes.

Treatment advice: Fluid Restriction: This is the first-line treatment for SIADH, aiming to reduce water retention. **Address the Underlying Cause:** If a malignancy or other underlying pathology is found (e.g., lung cancer), treating it may resolve the SIADH.

Salt Tablets/Hypertonic Saline: If the patient is symptomatic from hyponatremia or her sodium continues to drop, salt tablets or, in severe cases, slow infusion of 3% hypertonic saline may be necessary. Tolvaptan (vasopressin receptor antagonist) can be considered if fluid restriction is insufficient or impractical, though it should be used with caution, especially in elderly patients.

Comment

In 2023 ChatGPT gave no remark concerning the hyponatremia. In 2024 ChatGPT diagnosed SIADH instead of Addison's disease, and therefore hazardous water restriction was recommended. Adrenal insufficiency was wrongfully ruled out because of the normal potassium and creatinine.

Case 2 [1]

A 21-year-old woman with myasthenia gravis suffered from complaints of progressive weakness and fatigue over the past 6 months. Her appetite was poor, resulting in 2.5 kg weight loss. Plasma sodium was 121 mmol/L, potassium 7.6 mmol/L, glucose 3.2 mmol/L, blood pressure 73/44 mmHg, heart rate 118. Plasma creatinine was 4.8 mg/dl (423,42 μ mol/ml). During the initial 14 h, she received 4.8 L isotonic saline and excreted 4.7 L urine with an Uosm of 442 mOsm/kg H₂O and a urine sodium of 41 mmol/L and urine potassium 42 mmol/L [1].

Hoorn et al.: [1] Because of myasthenia gravis, chronic complaints, high potassium and low blood pressure, Addison's disease based on auto-immune disease was part of the diagnosis. However, crucial in this case was an increased risk of iatrogenic excessive increase of plasma sodium. This is because hypocortisolism was not the sole reason for hyponatremia as the patient was catabolic. The low osmol induced chronic hyponatremia is a risk factor for osmotic demyelination syndrome (ODS) as treatment with infusion of normal saline or hypertonic saline may cause an overly rapid rise in serum sodium [1, 5, 7].

Hoorn et al. [1] illustrated that once hyponatremia is associated with hypovolemia and decreased renal function, the fractional excretion of sodium (FE_{Na}) shows if the filtered load of sodium is much lower than serum

sodium tests insinuate. A $FE_{Na} < 1\%$ presumes prerenal azotemia as the kidney is avidly conserving sodium to maintain an effective circulating volume. In this case, the plasma creatinine was 6 times higher, reducing the GFR to about 20 ml/min (1/6 of the normal 120 ml/min). The filtered sodium load should then be 2420 μ mol/min (plasma sodium of 121 μ mol/ml x GFR of 20 ml/min). In this case the Na + K excretion in 12 h was 80 mmol/L in 4.5 L, thus 360 μ mol in 12 h and about 720 μ mol/day (500 μ mol/min), while the usual daily excretion of sodium is about 160 mmol/day and 60 mmol/day potassium [22]. Because of the high sodium and potassium excretion, another factor besides Addison's disease should be responsible for the high sodium loss. In other words, if hypovolemia had been the sole reason for the sodium decrease, we would expect a fractional excretion of sodium below 1%, but in this case, it was about 25% (500/2420 μ mol/min), suggesting that a second 'renal lesion' contributed to the excessive excretion of sodium.

The patient had a 6-month history with poor appetite and weight loss, signifying a low solute intake [1]. Examples of a low solute diet are "beer potomania" and "tea and toast diet" [23, 24, 25]. Under these circumstances a dangerous increase of the plasma sodium concentration may follow administration of solutes and fluids such as NaCl 0.9% [1, 5, 7]. The sudden availability of solutes and water amidst ADH depression will lead to massive free water diuresis. Normal saline will lower ADH, initiating water loss due to a high volume of osmol free urine, causing a rapid rise of plasma sodium. Important clues for a low salt and protein intake are a low blood urea nitrogen because of the low protein diet, and a urine osmolality below the serum osmolality. Brisk diuresis shortly after therapy also should warn physicians of a possible unwanted rise in plasma sodium [23, 24, 25].

The 46 physicians. *The crucial low osmol intake with increased risk of ODS was missed, but adrenal insufficiency was diagnosed by 19%. Inadequate fluid treatment with increased risk of cerebral damage was recommended by 57% of clinicians [1].*

ChatGPT (2023): *Given the electrolyte abnormalities, particularly the low sodium and high potassium levels, along with the elevated plasma creatinine, it suggests a potential adrenal crisis. This could be related to adrenal insufficiency, which may occur in the setting of myasthenia gravis or as a separate condition. The initial treatment involves fluid resuscitation with isotonic saline. The next step includes corticoid replacement.*

ChatGPT (2024): *This patient likely has primary adrenal insufficiency (Addison's disease). Immediate management should focus on stabilizing her through hormone replacement (hydrocortisone), fluid resuscitation, and addressing electrolyte imbalances. Long-term*

management will involve lifelong glucocorticoid and mineralocorticoid replacement.

Comment

Hypocortisolism related to the autoimmune disease is only one of two correct diagnoses. The important aspect of the low osmol intake was overlooked, and infusion of isotonic saline increased the plasma sodium rapidly to 129 mmol/l causing an osmotic demyelination syndrome (ODS) with a vegetative state [1, 20]. This second contributing cause could not only be derived from the low solute intake history, but also by evaluating the urine electrolytes. After therapy with NaCL 0.9%, the urine Na + K concentration (80 mmol/L) was much lower than the serum sodium concentration (121 mmol/L). This clearly showed the presence/ ability for free water clearance with a high risk of developing ODS due to a rapid and excessive increase of serum sodium, induced by infusion of NaCL 0.9% and ensuing free water excretion [23]. Acute kidney injury (AKI) was possibly an additional factor for the development of ODS, because AKI disturbs the function of the kidney to dilute or concentrate urine. The low blood pressure with reduced GFR will decrease the filtered load of sodium, but sodium reabsorption will increase in the proximal tubule where aldosterone does not act, further increasing the plasma sodium concentration. The possible unsafe use of normal saline should be emphasized, because chronic hyponatremia in combination with a low osmol intake increases the risk of ODS [7, 9]. Therefore, the plasma sodium should not increase more than 4 to 6 mmol/l in the first 24 h. In this case, half-isotonic saline would be safer to prevent a rapid increase of plasma sodium [1, 24].

ChatGPT correctly diagnosed Addison's disease in 2023 and 2024 but missed the important additional diagnosis of low sodium and protein intake and thus recommended the unsafe use of isotonic saline.

Case 3 [1]

A 31-year-old woman jogged 6 miles per day. She used to drink several liters of water per day. She was vegan and had a low salt intake. No history of alcohol, or drugs use. She urinated a large volume each day. The physical examination was normal, except that the body weight was low. The plasma sodium was 130 to 132 mmol/L on several occasions with a plasma osmol of about 269 mOsm/kg H₂O, urine osmol 79 mOsm/kg H₂O, and urine sodium was 11 mmol/l. Plasma creatinine was 0.93 mg/dl (82.23 µmol/ml); plasma blood urea nitrogen 5.9 mg/dl (2.11 mmol/L); Plasma potassium was 4.1 mmol/l. Thyroid and cortisol tests were normal [1].

As discussed above, Hoorn et al. [1] commented that subjects with an adequate osmol intake can drink about 18-liters of water per day without developing

hyponatremia. Excessive water intake during endurance sports may cause hyponatremia [7, 23, 24], but in this case the patient's diet with a low solute and salt intake in combination with a relative excess water intake seemed more important as the main hyponatremia cause. Infusion of normal saline or balanced solutions may cause a rapid and dangerous high free water clearance during low ADH action revealed by the low urine osmolality [1].

The 46 physicians. *The clinicians missed the low osmol intake as a the most important conclusion and their incorrect diagnosis included psychogenic polydipsia in 53% and SIADH in 12%. Adequate hypotonic solution treatment was recommended by 2% of the clinicians [1].*

ChatGPT (2023): *The combination of excessive water intake, long-distance jogging, and low salt intake raises the suspicion of exercise-associated hyponatremia (EAH). EAH can occur when an individual loses both water and sodium through sweating but replaces the losses with large volumes of water alone, leading to a dilution of sodium levels in the blood. The low urine osmolality (79 mOsm/kg H₂O) and low urine sodium (11 mmol/l) suggest an inability to concentrate urine appropriately, which can be a characteristic feature of EAH. In this context, the kidneys are not retaining sodium effectively.*

Treatment Proposal: The management of exercise-associated hyponatremia involves rehydration with a balanced solution that includes electrolytes. It's important to emphasize the importance of not only water but also replenishing sodium lost through sweat during prolonged exercise. Restricting excessive water intake during exercise and ensuring adequate salt intake can be preventive measures.

ChatGPT (2024): *Urine osmolality of 79 mOsm/kg suggests the kidneys are excreting very dilute urine, meaning there's an excessive intake of free water relative to solute. The patient's vegan, low-salt diet and high water intake could be contributing to primary polydipsia or exercise-associated hyponatremia, both of which can lead to dilutional hyponatremia. Urine sodium of 11 mmol/L supports the idea of sodium depletion or low sodium intake. Given the patient's low sodium diet, she is likely not compensating for sodium losses through sweat during jogging in a hot tropical environment.*

Conclusion: Exercise-Associated Hyponatremia or Primary Polydipsia: This diagnosis is consistent with chronic overconsumption of free water, especially in a hot climate with significant sweat loss and low sodium replacement.

Treatment Proposal: mild fluid restriction: encourage the patient to reduce water intake to below 2 L per day. This will help limit the dilution of sodium in the blood. Increase sodium Intake: Advise the patient to increase dietary salt intake to help replace sodium lost in sweat during exercise. Electrolyte solutions: Encourage the use

of electrolyte-rich beverages (not just plain water) during exercise to replace both water and sodium losses.

Comment

In 2023 ChatGPT missed the diagnosis low-osmol diet with high water intake and recommended an inadequate therapy with “balanced solution that includes electrolytes”. In 2024 the correct diagnosis was suggested by ChatGPT initially but later the conclusion switched to primary polydipsia or exercise-associated hyponatremia.

Case 4 [1]

A 21-year-old female became unwell at a party. On arrival at the emergency ward, she had a grand mal seizure. In blood drawn within 5 min after the seizure, her plasma sodium was 130 mmol/l. Body temperature was 39.8 °C; other vital and hemodynamic signs were normal. Her friend reported ecstasy intake, dancing and drinking water.

Hoorn et al.: [1] Ecstasy and water intake caused a severe hyponatremia that was missed because the sodium level increased temporarily after seizures [23, 26]. An important clue in this case is that severe symptoms such as seizures, related to acute hyponatremia (i.e. within 48 h), usually only occurs if the plasma sodium decreases substantially below 135 mmol/L, and in fact generally below 125 mmol/l [5, 7, 9]. The central aspect for the correct diagnosis in this case is that seizures can raise the plasma sodium acutely by 10–15 mmol/l [1, 20, 26]. This is the result of new osmoles, generated in the skeletal muscle cells, that attract water to shift from the extracellular fluid to the intracellular compartment, temporarily raising the sodium level. Reevaluation of plasma sodium within one to two hours is thus imperative to determine the steady state plasma sodium value [5, 6, 7, 8, 9].

The 46 physicians. *None of the clinicians addressed the acute effect of seizures on the sodium concentration. The clinicians incorrectly diagnosed SIADH, in 11%, water ingestion, in 13% and sodium loss in sweat by 28%. Proper administration of hypertonic saline was advised by 24% of the clinicians.*

ChatGPT (2023): *Hyponatremia due to the combination of ecstasy use, dancing, and drinking water. The seizure and elevated body temperature are consistent with the neurological effects of hyponatremia. The treatment for hyponatremia typically involves fluid restriction and, in severe cases, administration of hypertonic saline to correct the electrolyte imbalance.*

ChatGPT (2024): *The most likely diagnosis is 3,4-Methylenedioxymethamphetamine (MDMA) or “Ecstasy” intoxication, complicated by mild hyponatremia and hyperthermia.*

Treatment Proposal: Seizure management: If the seizure persists or reoccurs, administer a benzodiazepine. Given

the slightly reduced sodium level, fluid restriction is essential to prevent worsening of the dilutional hyponatremia. The patient should be closely monitored for changes in sodium levels. Hypertonic saline (3% NaCl) may be considered if the patient shows signs of worsening hyponatremia or severe neurological symptoms.

Comment

ChatGPT nor the clinicians addressed the acute effect of seizures on the sodium concentration. The advice of fluid restriction by ChatGPT in 2023 as well as 2024 was inappropriate. Proper administration of hypertonic saline was suggested by ChatGPT in 2024, but only after worsening of hyponatremia or neurological signs.

Discussion

ChatGPT has a great potential for helping healthcare professionals in complex diagnostic and therapeutic reasoning, but presently there are scant data on its diagnostic accuracy of difficult clinical cases. In this analysis of clinical vignettes from published case reports, ChatGPT 3.5 was not able to give a correct diagnosis with therapy plan in any of the four challenging hyponatremia cases in 2023 and only one out of four main aspects of the complex cases were diagnosed properly when used in 2024. Forty-six physicians from several countries had similar difficulties in diagnosing and managing these cases [1]. ChatGPT referring to its limitations, states that “given the potential seriousness of hyponatremia, decisions about diagnosis and treatment should be made in consultation with healthcare professionals who have access to the patient’s complete medical history and can consider additional diagnostic tests if needed and provide personalized advice on fluid and electrolyte balance”. However, physicians similarly struggle with the differential diagnosis of hyponatremia.

The evaluation of hyponatremia always starts with a detailed history and physical examination [20, 23, 24, 27, 28, 29], and as history and physical examination are notoriously inaccurate to determine a patient’s volume status [4, 6, 9], inaccuracies can easily misguide both artificial intelligence and clinicians. In addition, common flowcharts are often centralized on this difficult to assess volume status and are less applicable when there are multifactorial causes of hyponatremia.

Artificial intelligence has the capacity to deal with a huge set of medical facts and because of the ability to learn from new data and medical advances, Chat GPT may reduce diagnostic errors, also in rare diseases [30, 31, 32, 33, 34]. Nevertheless, one cannot fully trust its judgment for medical cases yet, also because ChatGPT may produce (‘hallucinate’) details that contain incorrect facts, leading to mistakes [34]. Medical diagnosis using artificial intelligence will undoubtedly in due time

become a relevant healthcare decision support system. However, ChatGPT acquires data from the current databases and potential errors of the technology may be the result of inaccurate information accumulated in the literature over time as well as from the electronic health record [31, 34].

There are several limitations in this study. Only four cases were evaluated, merely related to hyponatremia. The case vignettes are uncommon and do not represent most everyday patient presentations, but even then, physicians should be knowledgeable about the aspects of these cases, as errors in diagnosis and therapy can have devastating effects. Perhaps simple hyponatremia cases would have been correctly diagnosed by both physicians and ChatGPT, but in clinical practice, cases are seldomly simple, and insufficient knowledge and lack of proper testing and tests interpretation may lead to erroneous conclusions [23]. ChatGPT is an evolving program and improvements may follow regularly, as shown by some progress of the case management in 2024 as compared to 2023. In the scenario described in this paper complex cases were evaluated in which physicians also struggled with a proper assessment, despite the use of current algorithms [1]. Moreover, the used cases predated 2021, and therefore may have been included in the initial training set for ChatGPT. We do not know if ChatGPT is copying old information or generating a unique response to our prompts but considering the low improvement of the same cases after 10 months, this learning effect seems rather low.

Notwithstanding these limitations, patient cases studies may improve AI knowledge regarding clinical diagnosis, provided all (potential) pitfalls are presented [35, 36, 37, 38]. However, in a recent study of 150 Medscape cases that were presented to ChatGPT 3.5, only 49% were answered correctly [34].

Compared with the free for use ChatGPT 3.5, the large language model (LLM), ChatGPT 4.0 demanding a paid subscription, has improved accuracy when answering certain medical multiple-choice questions [13] but it remains to be seen if its use can overcome some of the current shortcomings of ChatGPT 3.5. Presently, ChatGPT-4 is not specifically designed for medical tasks [12], but considering the speed of learning, reasoning, and self-correction by artificial intelligence, rapid progress in diagnosing complex cases can be anticipated. In a recent study ChatGPT-4 provided the correct diagnosis in 64% of challenging cases, using the New England Journal of Medicine clinicopathologic conferences [35]. However, in another recent trial, the availability of ChatGPT-4 to physicians as a diagnostic aid did not significantly improve clinical reasoning compared with conventional resources such as UpToDate and Google [36].

Nevertheless, concerns such as accuracy, safety, validity, ethical implications and issues on data protection and privacy remain before full integration of artificial intelligence for diagnostic purposes in healthcare [12–14, 30–37]. Furthermore, as shown for these difficult hyponatremia cases the ChatGPT-3.5 platform needs the input of specialists providing the medical data and the necessary information to its database and they must themselves be well trained to correctly address these complex cases. Training of frequently encountered difficult medical subjects should be, in the future, part of medical student and resident training, if possible aided by artificial intelligence platforms [32, 33, 34, 35, 36, 37, 38].

Conclusion

The use of ChatGPT 3.5 poses a risk of errors in diagnosis and therapy when evaluating certain complex cases of hyponatremia. Health professionals, patients and people without medical training may have an unjustified trust on diagnosis and therapy suggestions by ChatGPT. Therefore, also students should be aware of the potential shortcomings when using AI support for addressing patient evaluation tasks [36]. Current artificial intelligence diagnostic shortcomings probably will improve soon, but in the meantime critical evaluation remains crucial. One way to evaluate the development is to perform complex cases assessment on a regular basis for system validation.

Abbreviations

| | |
|------------------|--|
| ChatGPT | Chat Generative Pre-trained Transformer |
| CDA | Clinical diagnostic algorithms |
| ACTH | Adrenocorticotrophic hormone |
| ADH | Antidiuretic hormone |
| SIADH | Syndrome of inappropriate ADH secretion |
| ECF | Extracellular fluid or extracellular volume |
| ODS | Osmotic demyelination syndrome |
| FE _{Na} | Fractional excretion of sodium |
| EAH | Exercise-associated hyponatremia |
| MDMA | 3,4-Methylenedioxymethamphetamine or "ecstasy" |
| LLM | Large language model |

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12909-025-07235-2>.

Supplementary Material 1

Acknowledgements

Ibrahim Ajubi performed technical support. Zhoreh Safipour performed technical support.

Author contributions

K.B. designed the study and drafted the first version of the article. All authors contributed to the content and reviewed the manuscript.

Funding

None.

Data availability

Availability of data and materials: provided in a supplementary file.

Declarations

Ethics approval and consent to participate

Ethical framework: because the four anonymous case reports adapted for this study were previously published and are openly accessible, no approval from our Medical Ethics Committee was required and the study was exempt from the Declaration of Helsinki.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Medicine, Curaçao Medical Center, Willemstad, Curaçao

²Institute for Medical Education, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

³Red Cross Blood Bank Foundation, Willemstad, Curaçao

⁴Department of Medicine, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

Received: 1 November 2024 / Accepted: 25 April 2025

Published online: 22 May 2025

References

- Hoorn EJ, Halperin ML, Zietse R. Diagnostic approach to a patient with hyponatraemia: traditional versus physiology-based options. *QJM*. 2005;98(7):529–40.
- Boscoe A, Paramore C, Verbalis JG. Cost of illness of hyponatremia in the United States cost effectiveness and resource allocation 2006, 4:10 <https://doi.org/10.1186/1478-7547-4-10>
- Buckenmayer A, Siebler N, Haas CS. Evaluation of simple diagnostic parameters in acute kidney injury in hospitalized patients—diagnostic recommendations for non-nephrologists. *Intern Emerg Med*. 2023;18:1769–76. <https://doi.org/10.1007/s11739-023-03365-x>.
- Soiza RLRL, Kirsten Cumming K, Clarke JM, Wood KM, et al. Hyponatremia: special considerations in older patients. *clin. Med*. 2014;3:944–58. <https://doi.org/10.3390/jcm3030944>.
- Arnaldi G, Arvat E, Berton AM, et al. Endocrinologists at work: management of hyponatremia in clinical practice. *J Endocrinol Invest*. 2023;46:2543–7. <https://doi.org/10.1007/s40618-023-02147-8>.
- McGee S, Abernethy WB 3rd, Simel DL. The rational clinical examination. Is this patient hypovolemic? *JAMA*. 1999;17(11):1022–9. <https://doi.org/10.1001/jama.281.11.1022>.
- Verbalis JG, Goldsmith SR, Greenberg A, et al. Diagnosis, evaluation, and treatment of hyponatremia: expert panel recommendations. *Am J Med*. 2013;126(10 Suppl 1):S1–42. <https://doi.org/10.1016/j.amjmed.2013.07.006>.
- Musch W, Decaux G. Utility and limitations of biochemical parameters in the evaluation of hyponatremia in the elderly. *Int Urol Nephrol*. 2001;32:475–93.
- Hoorn EJ, Zietse R. Diagnosis and treatment of hyponatremia: compilation of the guidelines. *J Am Soc Nephrol*. 2017;28(5):1340–9.
- Fenske W, Maier SK, Blechschmidt A, Allolio B, Störk S. Utility and limitations of the traditional diagnostic approach to hyponatremia: A diagnostic study. *Am J Med*. 2010;123:652–7.
- Upadhyay A, Jaber BL, Madias NE. Incidence and prevalence of hyponatremia. *Am J Med*. 2006;119:S30–5.
- A. openAI official website [online]. Available from: <https://openai.com/about>; Jan 2024).
- Eriksen AV, Møller S, Ryg J. Use of GPT-4 to diagnose complex clinical cases. *NEJM AI*. 2023;1(1). <https://doi.org/10.1056/AI2300031>.
- Shea Y, Min Yao Lee C, Chin Tung Ip W, et al. Use of GPT-4 to analyze medical records of patients with extensive investigations and delayed diagnosis. *JAMA Netw Open*. 2023;6(8):e2325000. <https://doi.org/10.1001/jamanetworkopen.2023.25000>. (Reprinted).
- Niel O, Bastard P. Artificial Intelligence in Nephrology: Core Concepts, Clinical Applications, and Perspectives. *Am J Kidney Dis*. 2019;74(6):803–810. <https://doi.org/10.1053/j.ajkd.2019.05.020>. Epub 2019.
- Wang L, Li W. The impact of AI usage on university students' willingness for autonomous learning. *Behav Sci*. 2024;14:956. <https://doi.org/10.3390/bs14100956>.
- Voglis S, van Niftrik CHB, Staartjes VE, Brandi G, Tschopp O, et al. Feasibility of machine learning based predictive modelling of postoperative hyponatremia after pituitary surgery. *Pituitary*. 2020;23:543–51.
- Theerthagiri P. Forecasting hyponatremia in hospitalized patients using multilayer perceptron and multivariate linear regression techniques. *Concurrency Computat Pract Exper*. 2021;33:e6248.
- Kinoshita T, Oyama S, Hagiwara D, et al. A machine learning approach for predicting treatment response of hyponatremia. *Endocr J*. 2024;71(4):345–55. <https://doi.org/10.1507/endocrj.EJ23-0561>.
- Van Steijn JH, Gans RO. Clinical thinking and decision making in daily practice. An elderly man with hyponatraemia. *Ned Tijdschr Geneesk*. 2002;146:404–10.
- Weidmann P, Beretta-Piccoli C, Glück Z, Keusch G, Reubi FC, De Châtel R, et al. Hypoadosteronism without hyperkalemia. *Klin Wochenschr*. 1980;58:185–94. <https://doi.org/10.1007/BF01476777>.
- Valero-Morales I, Tan M, Pei Y, He FJ, MacGregor GA. 24-hour sodium and potassium excretion in the Americas: a systematic review and meta-analysis. *Rev Panam Salud Publica*. 2022;46:e199. <https://doi.org/10.26633/RPSP.2022.199>.
- Berend K, Heldeweg M. The role of the clinical laboratory in diagnosing hyponatremia disorders. *Crit Rev Clin Lab Sci*. 2025. Mar;1:1–26. Online ahead of print. PMID: 40022565 Review.
- Spasovski G, Vanholder R, Allolio B, et al. Clinical practice guideline on diagnosis and treatment of hyponatraemia. *Eur J Endocrinol*. 2014;170:G1–47. <https://doi.org/10.1530/EJE->.
- Welt LG, Orloff J, Kydd DM, Oltman JE. An example of cellular hyperosmolarity. *J Clin Invest*. 1950;29:935–9.
- Watson R. Ecstasy is making a comeback, European drugs report warns. *BMJ*. 2016; 353 <https://doi.org/10.1136/bmj.i3078> BMJ 2016;353:i3078.
- Theodosios D, Filippatos GL, Moses S, Elisaf. Ten pitfalls in the proper management of patients with hyponatremia. *Postgrad Med*. 2016;128:5516–522. <http://doi.org/10.1080/00325481.2016.1186488>.
- Hoorn EJ, Hotho D, Hassing RJ, Zietse R. Unexplained hyponatremia: seek and you will find. *Nephron Physiol*. 2011;118:p66–71.
- Nagarur A, Axelrod L, Dighe AS. Case 9-2017: A 27-Year-Old woman with nausea, vomiting, confusion, and hyponatremia. *N Engl J Med*. 2017;376:1159–67. <https://doi.org/10.1056/NEJMcpc1616024>.
- Ruksakulpiwat S, Kumar A, Ajibade A. Using ChatGPT in medical research: current status and future directions. *J Multidiscip Healthc*. 2023;16:1513–20. <https://doi.org/10.2147/JMDH.S413470>.
- Kharat PB, Dash KS, Rajpurohit L, Tripathy S, et al. Revolutionizing healthcare through chat GPT: AI is accelerating medical diagnosis. *Oral Oncol Rep*. 2024;9:100222.
- Shieh A, Tran B, He G, et al. Assessing ChatGPT 4.0's test performance and clinical diagnostic accuracy on USMLE STEP 2 CK and clinical case reports. *Sci Rep*. 2024;14:9330. <https://doi.org/10.1038/s41598-024-58760-x>.
- Khanjee Z, Crowe B, Rodman A. Accuracy of a generative artificial intelligence model in a complex diagnostic challenge. *JAMA*. 2023;330(1):78–80. <https://doi.org/10.1001/jama.2023.8288>.
- Hadi A, Tran E, Nagarajan B, Kirpalani A. Evaluation of ChatGPT as a diagnostic tool for medical learners and clinicians. *PLOS ONE*. 2024;19(7):e0307383. <https://doi.org/10.1371/journal.pone.0307383>.
- Tenner ZM, Cottone MC, Chavez MR. Harnessing the open access version of ChatGPT for enhanced clinical opinions. *PLOS Digit Health*. 2024;3(2):e0000355. <https://doi.org/10.1371/journal.pdig.0000355>.
- Goh E, Gallo R, Hom J, et al. Large Language model influence on diagnostic reasoning A randomized clinical trial. *JAMA Netw Open*. 2024;7(10):e2440969. <https://doi.org/10.1001/jamanetworkopen.2024.40969>.
- Shalong W, Yi Z, Bin Z et al. Enhancing self-directed learning with custom GPT AI facilitation among medical students: A randomized controlled trial. *Med Teach*. 2024. doi: 10.1080/0142159X.2024.2413023. Online ahead of print.
- Chloe Antoniou AP, Danagra G, Ikossi. 31 Oct (2024): Let's chat! Integrating ChatGPT in medical student assignments to enhance critical analysis, Medical Teacher. <https://doi.org/10.1080/0142159X.2024.2421997>

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.