nature medicine

Article

https://doi.org/10.1038/s41591-025-03601-1

AI-guided precision parenteral nutrition for neonatal intensive care units

Received: 23 October 2024

Accepted: 17 February 2025

Published online: 25 March 2025

Check for updates

A list of authors and their affiliations appears at the end of the paper

One in ten neonates are admitted to neonatal intensive care units, highlighting the need for precise interventions. However, the application of artificial intelligence (AI) in guiding neonatal care remains underexplored. Total parenteral nutrition (TPN) is a life-saving treatment for preterm neonates; however, implementation of the therapy in its current form is subjective, error-prone and resource-consuming. Here, we developed TPN2.0—a data-driven approach that optimizes and standardizes TPN using information collected routinely in electronic health records. We assembled a decade of TPN compositions (79,790 orders; 5,913 patients) at Stanford to train TPN2.0. In addition to internal validation, we also validated our model in an external cohort (63,273 orders; 3,417 patients) from a second hospital. Our algorithm identified 15 TPN formulas that can enable a precision-medicine approach (Pearson's $R = 0.94$ compared to experts), increasing safety and potentially reducing cost. A blinded study ($n = 192$) revealed that physicians rated TPN2.0 higher than current best practice. In patients with high disagreement between the actual prescriptions and TPN2.0, standard prescriptions were associated with increased morbidities (for example, odds ratio = 3.33; P value = 0.0007 for necrotizing enterocolitis), while TPN2.0 recommendations were linked to reduced risk. Finally, we demonstrated that TPN2.0 employing a transformer architecture enabled guideline-adhering, physician-in-the-loop recommendations that allow collaboration between the care team and AI.

Neonatal intensive care units (NICUs) provide critical care to newborns with serious medical conditions. In the United States, approximately 10% of newborns are admitted to NICUs and this number continues to rise[1–3]. Preterm birth—the leading cause of death in children under five years old—is a substantial contributor to NICU admissions, with its rate also increasing over the past decade[4]. To better serve these vulnerable newborns, NICU practices have incorporated increasingly advanced innovations. These range from the introduction of mechanical ventilation in the 1970s[5], the implementation of surfactant replacement therapy in the 1980s[6] and the use of therapeutic hypothermia in the United States in early 2000s[7]. Today, with the rise of big data and AI, another new frontier in NICU innovation involves making care decisions informed by computational methods that leverage troves of data, such as continuous monitor recordings and electronic health records (EHRs). We are already seeing advances in various predictive models, such as for the detection of newborn sepsis[7] or intraventricular hemorrhage (IVH)[8]. Despite these advances, the potential of AI remains largely untapped. Clinical adoption of AI in many US hospitals remains as low as 10% (ref. 9), with almost 90% of the adoption concentrated in radiology and cardiology and 0% in neonatology[10]. Even with high accuracy, many predictive models do not influence outcomes in real clinical settings[11,12]. This underscores the need for clinically relevant AI that goes beyond making predictions to guiding effective interventions, a shift reflected in emerging trends such as large action models in autonomous systems, which emphasize actionable outcomes for greater clinical impact[13,14].

e-mail: naghaeep@stanford.edu

TPN, used here to encompass both TPN (zero enteral feeds) and partial PN (parenteral nutrition), where some amounts of enteral feeds are also provided, plays a crucial role for many NICU patients. It is in particularly important for those born prematurely or with gastrointestinal complications, by providing essential nutrients directly into the bloodstream when enteral feeding is not possible or sufficient.

For some newborns with underdeveloped or compromised gastrointestinal tracts, TPN is their main source of nutrition and sole source for some[15]. However, best practice for prescribing and formulating TPN is costly, time-consuming to formulate and requires extensive multidisciplinary collaboration. Other practices that do not involve a multidisciplinary team or adhere to standardized guidelines tend to result in worse outcomes[16,17]. Contributing to its complexity are the regional differences in practice and the lack of universal guidelines[16,18]. In the United States, TPN protocols for preterm infants emphasize early, aggressive nutrition with higher protein and calorie goals; in Europe, formulations with lower proteins are used to avoid metabolic complications[19–21]. Given these complexities, it is not surprising that errors in TPN management—ranging from prescribing and compounding to labeling and administration—are the most frequently cited among high-alert medications in NICUs[22,23]. To mitigate these errors, previous calculator tools for TPN formulations have aimed primarily at addressing issues such as component compatibility, electrolyte balance and osmolarity calculations[24–27]. These calculator tools are available commercially (for example, Infusion Studio or modules within Epic[28,29]. However, their scopes remain limited in addressing the broader complexities of prescribing TPN. Whether due to errors or suboptimal protocols, improper TPN administration has been associated with heightened risks of mortality and morbidities, including necrotizing enterocolitis (NEC), bronchopulmonary dysplasia (BPD), gastrointestinal diseases, sepsis and more[30–36]. These challenges and risks associated with TPN highlight an opportunity for data-driven approaches like AI to improve safety, efficiency and patient outcomes in the NICU.

Here, we report the development and validation of TPN2.0—an AI approach that formulates a set of standardized TPN compositions and assigns treatments to newborns based on routinely collected data from their EHR. We developed TPN2.0 using advanced machine learning (ML) algorithms, performed external validation and tested it with a multidisciplinary healthcare team. TPN2.0 can also have a broader impact beyond the standard of practice. The recent American Society for Parenteral and Enteral Nutrition (ASPEN) guidelines for parenteral nutrition in premature infants recognize that standardized solutions like TPN2.0, being relatively easy to implement, will be particularly valuable in resource-limited settings[20], including in the United States, and in low- and middle-income countries (LMICs) where customized TPN is often inaccessible[37,38].

Results

We compiled a TPN dataset consisting of 79,790 prescriptions from 5,913 unique patients and linked them to their EHRs to develop a data-driven TPN approach called TPN2.0. The data were collected at Stanford Health Care—a quaternary care hospital in the United States, from January 2011 to January 2022. The newborn clinical characteristics are reported in Supplementary Table 1 along with the incidence of 17 principal neonatal morbidities. The clinical characteristics collected include demographics, laboratory measurements, conditions, medications, observations and procedures. A flowchart detailing this process is shown in Extended Data Fig. 1. An independent dataset from the University of California, San Francisco (UCSF) was used for external validation of TPN2.0. The UCSF dataset contained 63,273 TPN prescriptions from 3,417 unique patients.

Study design and overview

TPN2.0 was designed to reduce the formulation subjectivity and increase the compounding efficiency of the TPN prescription process by leveraging ML to (1) formulate a set of standardized TPN formulas and then (2) recommend the best fit from these standardized formulas based on the patient's clinical characteristics (Fig. 1a). The approach is data-driven, resolving the ambiguity in determining the ideal composition of TPN, which currently varies based on selected guidelines and individual providers' anecdotal experiences[19,20,39]. This approach can streamline the prescription process, reducing the human resources and time required from hours to probably a couple of minutes. All of this is achieved while maintaining the necessary level of personalization to meet the diverse nutritional needs of newborns. The standardized formulas also open the possibility of mass manufacturing, which can eliminate the 4–12 h of lead time and errors from individual compounding and shipping. These benefits enable TPN2.0 to address the current TPN processes' shortcomings and accessibility issues.

TPN2.0 was developed based on data collected over a decade at Stanford (Supplementary Table 1). During this 10-year period, the patients' average number of days on TPN was 14 ± 23 days, with a median of 7 days reflecting high variability among cases; for example, the 90th percentile extends up to 50 days (Fig. 1b). The components of TPN can also vary independently over time (Fig. 1c and Supplementary Fig. 1). For example, all macronutrients monotonically increased and plateaued, whereas zinc and copper peaked and then decreased. This variability, alongside diverse patient phenotypes, underscores the complex challenges in nutritional management.

TPN2.0 formulas identified through a data-driven approach

To develop TPN2.0, we first employed a variant of deep-learning architecture called a variational neural network (VNN) combined with semisupervised iterative clustering (Fig. 2a). The VNN compressed high-dimensional patient data into a latent representation, which is effective for subsequent clustering. The TPN2.0 VNN was trained on the compressed EHR representations (Methods), daily laboratory values, demographic data and TPN basic parameters such as daily total fluid targets and enteral volumes, to generate the TPN compositions. The latent representation from the VNN was then grouped into discrete clusters to reduce the innumerable numbers of TPN recommendations into a finite set of clusters, that is, standardized formulas, that allows mass manufacturing. This could substantially reduce costs associated with TPN and increase safety. Moreover, standardized TPN can have a long shelf-life of up to 2 years at room temperature[40], making widespread distribution possible to hospitals in low-resource settings that currently lack means to independently formulate TPN.

We found that only 15 clusters were required for TPN2.0 to achieve sufficient correlations between a given cluster's composition and the prescribed TPN. This number of clusters yielded a Pearson's correlation (R) of 0.82 (P value < 0.0001) relative to the original VNN predictions. Two-dimensional (2D) visualization of the VNN's latent space demonstrates the heterogeneity of the nutritional needs in different patients (Fig. 2b and Supplementary Fig. 2). The optimal composition of TPN may change from day to day, and each patient may not necessarily follow the same longitudinal trajectory. An example is presented in Fig. 2b, where a patient requiring intravenous nutrition through a central line was first assigned to cluster C3, which contained the lowest concentrations of solutes (Fig. 2c). In response to higher nutritional needs as the newborn grew from days 2–9, they were then shifted from C3 to C1 and then C7, which provided higher nutritional and caloric content. However, upon being diagnosed with hyponatremia, TPN2.0 adapted by moving the patient to cluster C6, which delivered a higher sodium concentration of 5.6 mEq kg^{-1} day^{-1}. The nutritional profile of each cluster is shown in Extended Data Fig. 2. This flexibility highlights the capability of TPN2.0 to address diverse nutritional needs by algorithmically assigning one of the 15 standardized formulas.

The cluster analysis demonstrated that model performance using 15 clusters was optimal. A higher number of clusters resulted

https://doi.org/10.1038/s41591-025-03601-1

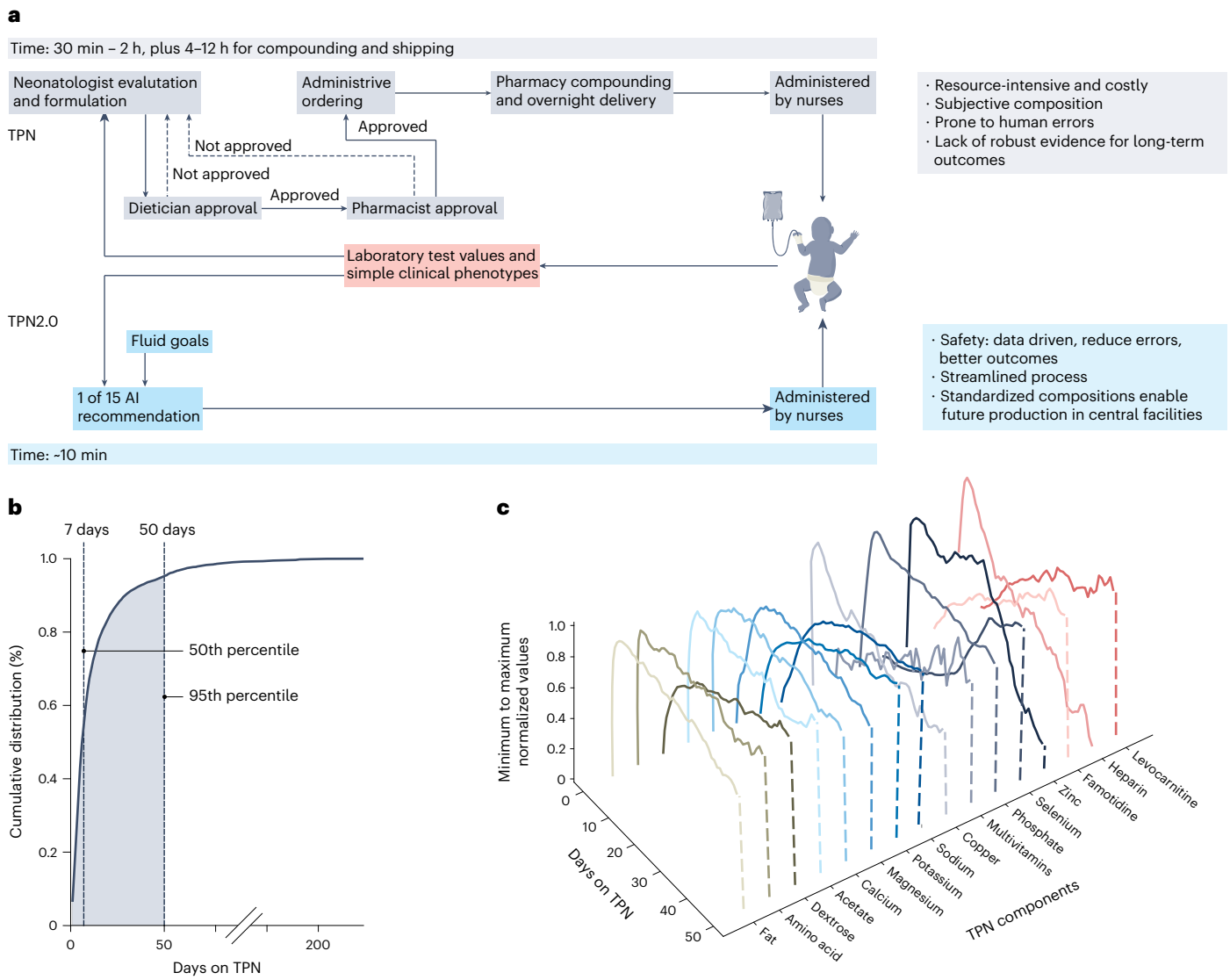


Fig. 1 | TPN ordering is a repetitive, time-consuming and error-prone process, with many stakeholders involved. a, The current TPN ordering workflow involves a multidisciplinary team collaborating to determine the appropriate daily TPN composition for each patient. A medical team first evaluates the patient's laboratory test results and clinical characteristics and place an order. The order is reviewed by a dietician and, if approved, by a pharmacist. Next, each bag is compounded individually, and delivered to the hospital. This process typically takes 4–12 h, and sometimes up to 24 h. In contrast, the proposed TPN2.0 model simplifies the process by automatically analyzing patient data—comprising laboratory test values and clinical characteristics—and assigning one of the 15 premade TPN formulas to them. TPN2.0 aims to streamline operations to reduce cost, error and practice variability. This is achieved by a combination of AI and standardized TPN compositions. **b**, Cumulative distribution of the number of days patients are on TPN shows that 50% of the patients receive TPN for up to 7 days, and 95% receive it for up to 50 days. **c**, Normalized mean values of TPN components by day, demonstrating the dynamic nature of TPN components.

in diminishing gains and did not show enough gain in performance to justify an increased number of formulas (Fig. 2d). TPN2.0 was also more objective than actual prescriptions. Our analysis showed that TPN2.0 had a lower variance for similar types of patients compared to current best practice across all TPN components (Extended Data Fig. 3). This indicates that TPN2.0 was more consistent than actual prescriptions in recommending the same TPN compositions for patients with similar clinical characteristics. It also demonstrates that any two members of the healthcare team may not agree on the same TPN compositions for the same patient (for more on this, see 'Blinded providers rate TPN2.0 higher than best practice').

Independent validation of TPN2.0 in a second hospital

In this retrospective study, we validated the VNN model on an independent cohort at UCSF. The Stanford team did not have access to the UCSF data. The model, without retraining, was provided to UCSF and the results were reported independently.

Although TPN practices may vary among providers, and more so across different hospital systems, certain characteristics remain similar. For example, the TPN dosage at both Stanford and UCSF can predict the weights of infants (Pearson's $R \geq 0.90$; Fig. 3a). At Stanford, TPN2.0 recommendations performance was highly correlated with that of human experts (Pearson's $R = 0.94$; P value < 0.0001; 20% difference) and vastly outperformed baseline ML methods such as Elastic Net (Fig. 3b). Difference between prescriptions among human experts for a similar type of patients was used as the gold standard (Methods). The resulting difference, measured by distance of each TPN component, among human experts and to TPN2.0 recommendations is shown in Fig. 3c. Stratification of the results indicated that the performance generalized similarly across races and sexes (Extended Data Fig. 4a). The high level of correlation persisted even when the train/test set in the

https://doi.org/10.1038/s41591-025-03601-1

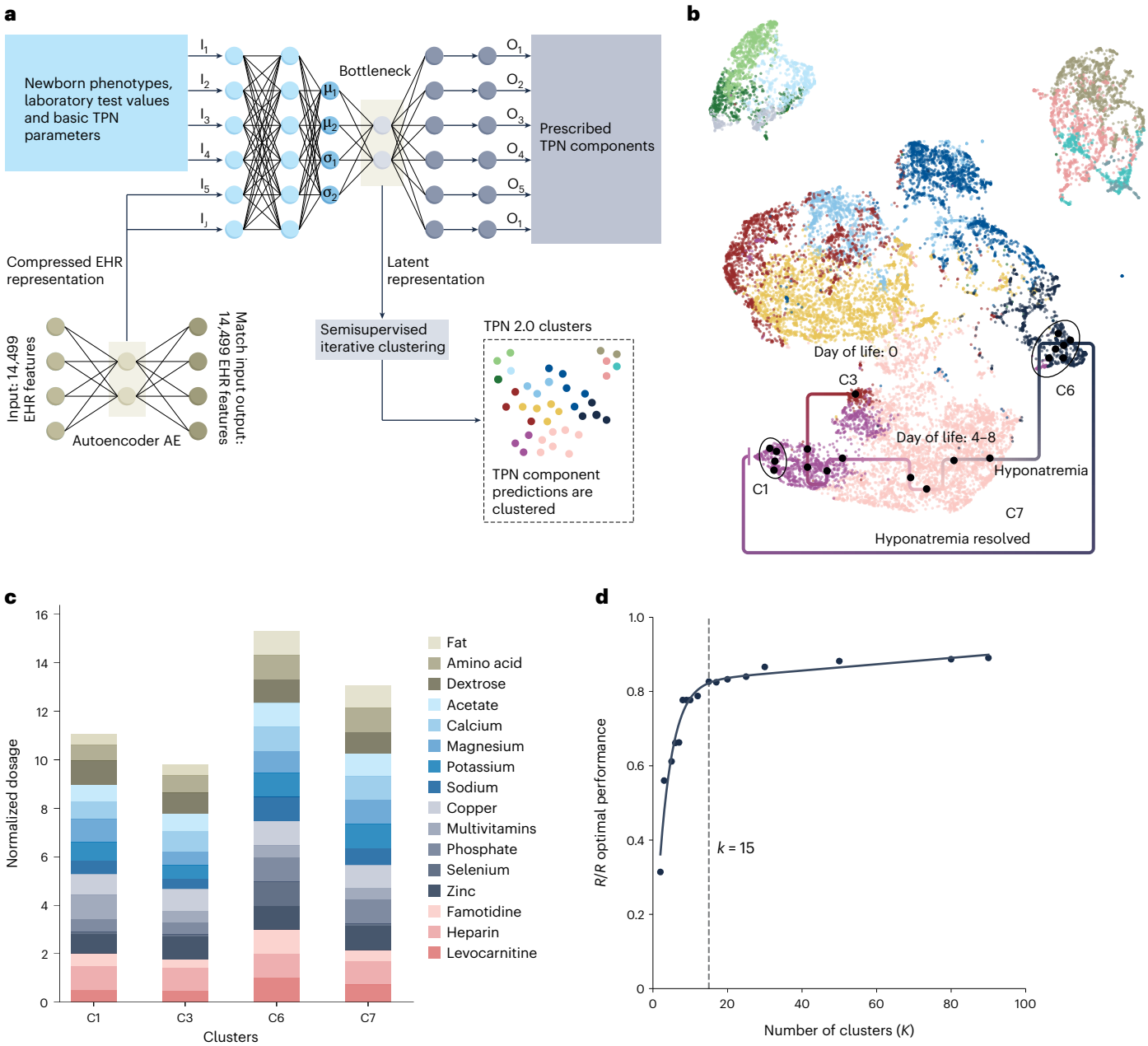


Fig. 2 | A deep representation learning algorithm for data-driven prediction and standardization of TPN. a, A VNN is used to predict TPN compositions based on patient data. The clinical characteristics include newborn characteristics, laboratory measurements and basic TPN information such as total fluid. The compressed EHR representations are obtained from the latent space of the AE fed with 14,499 EHR features including medications, observations, procedures and conditions. The VNN integrates all these inputs and maps them to 16 different TPN components, creating a high-dimensional latent representation that encompasses both prescription and patient information. These latent representations are then fed to a semisupervised iterative clustering algorithm to group similar representations together. The TPN composition of each cluster is obtained by feeding the cluster centroid to the VNN's decoder, yielding a set of standardized TPN2.0 formulas. **b**, A 2D visualization of the latent representations, color-coded by cluster assignment. Each dot represents a latent patient EHR profile of that day. The black dotted path corresponds to an example from a particular patient, transitioning through different TPN2.0 clusters based on their historical EHR. Initially, TPN2.0 recommends formula C3, followed by C1, which are low-concentration bags typically used in the first few days of life. Subsequently, the recommendation switches to a more nutrient-rich formula, C7, until hyponatremia develops. In response to this, the model automatically switches the patient to formula C6, which has one of the highest sodium concentrations. Once the hyponatremia is resolved, the recommendation goes back to C1. This case showcases how the model can adjust TPN bag prescriptions based on real-time changes in patient conditions. **c**, TPN2.0 clusters have distinct compositions. The within-nutrient normalized composition of the formulas illustrated in the example patient's journey. Comparison should not be made across nutrients. **d**, This plot visualizes the relative Pearson's *R* and the number of clusters, demonstrating diminishing returns after 15, where higher numbers of clusters did not result in substantial performance gain to justify the decrease in practicality. The relative *R* was calculated from the ratio between *R* from the cluster predictions and from the model's raw predictions.

time-based cross-validation scheme. Specifically, we trained the model on data from two timeperiods and tested it on the remaining period, repeating this process for each of the three periods. This approach demonstrated that the model's performance remained consistently high, regardless of which timeperiod was used for testing (Extended Data Fig. 4b). The stable correlation throughout all periods implies that

https://doi.org/10.1038/s41591-025-03601-1

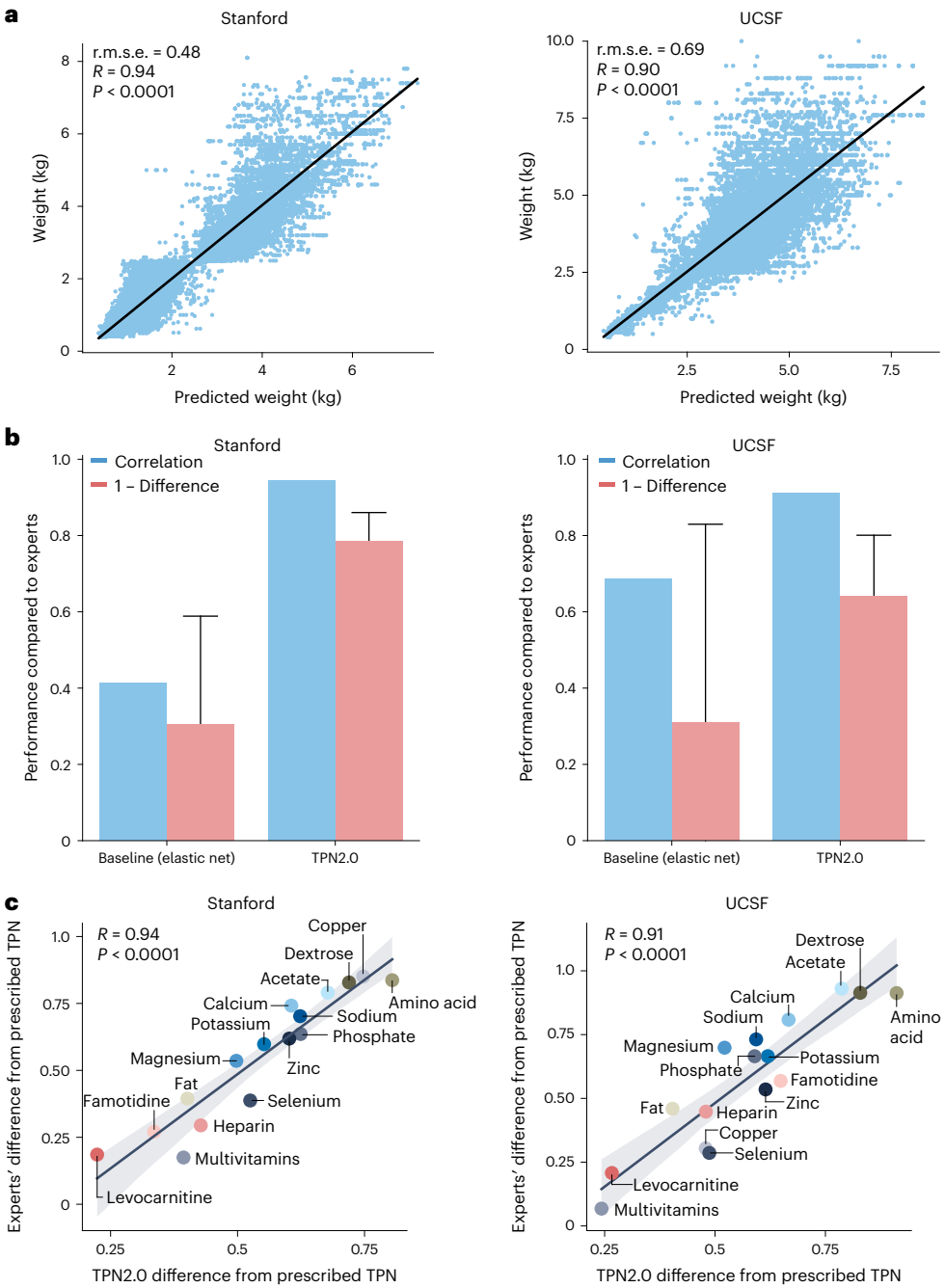


Fig. 3 | TPN2.0 is validated in a second hospital and outperforms baseline Elastic Net models. a, To validate the model developed at Stanford, we extracted a second TPN dataset from UCSF. To check data consistency, we demonstrated that TPN compositions are associated with the weights of neonates in both sites (Pearson's *R* and *P* value). **b**, The Stanford team did not have access to UCSF data, and models trained at Stanford were validated independently by the UCSF team. To examine TPN2.0 performance across sites, we report the distance between a prescribed TPN and another expert's similar TPN order as the gold standard (Methods). We also report the relative difference, which is calculated as the absolute difference between TPN2.0 and the experts' distances, divided by that of the expert. It is presented as '1 − Difference', where higher values signify higher similarity between experts and TPN2.0 performance. Error bars, s.e. Comparing these distances between the experts' TPN and TPN2.0 reveals a high correlation at both Stanford (n = 79,790 TPN from 5,913 patients) and UCSF (n = 63,273 TPN from 3,417 patients). The model also outperforms baseline ML (Elastic Net). Data are presented as mean values ± s.e.m. **c**, At both sites, TPN2.0 shows similar distance to experts for all components (Pearson's *R* and *P* value). Here, lower distances suggest consistency among experts, as seen with components like levocarnitine or multivitamins. In contrast, higher distances reflect lower performance, such as in dextrose and amino acid contents, which are due to greater variability in practice. In contrast, levocarnitine or multivitamins are mostly a binary decision with better defined guidelines. The bands represent 95% confidence intervals.

potential shifts in clinical guidelines over time had little influence on the effectiveness of the model.

At UCSF, TPN2.0 retained significantly high performance with their human experts (Pearson's R = 0.91; P value < 0.0001; 35% difference; Fig. 3b). At both sites, dextrose and amino acids were among the components exhibiting the highest distances for both the human experts and, to a lesser extent, TPN2.0 (Fig. 3c). This implies higher variability between expert providers in these components, which was expected since the guidelines for these components have wider possibilities. This is in contrast to those with lower distances such

https://doi.org/10.1038/s41591-025-03601-1

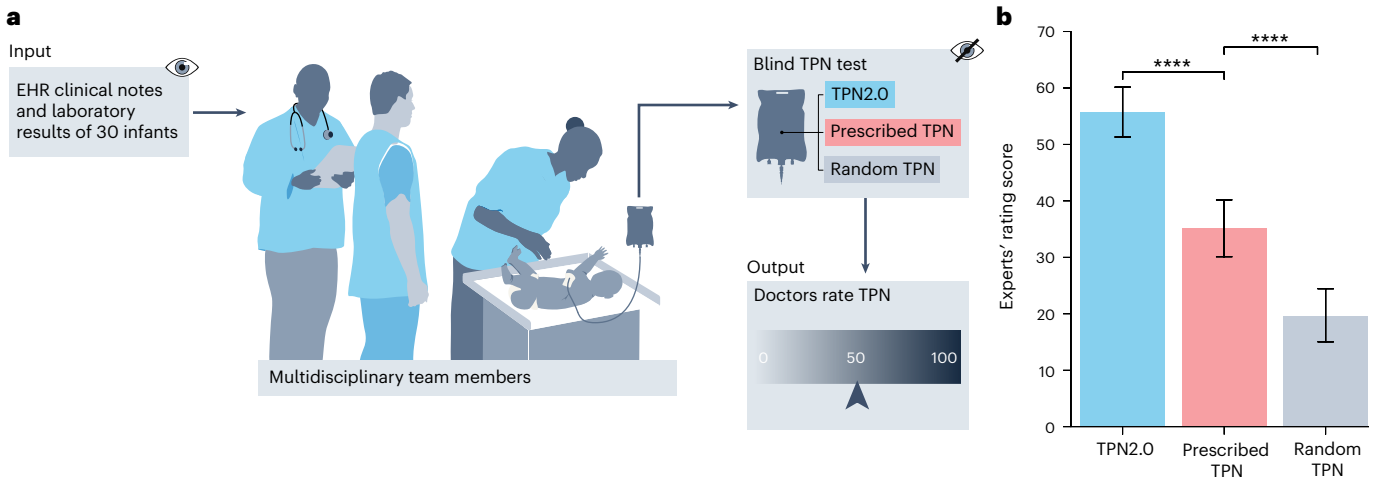


Fig. 4 | In a blinded study, TPN2.0 outperformed current best practice. **a**, In a blinded study, physicians who regularly prescribe TPN were recruited to rank three TPN solutions: TPN2.0, TPN composition from a different patient (randomly selected) and the actual prescription developed for that patient according to current best practice. Each team member ranked each of the three solutions from 0 to 100 after a thorough chart review based on all information available in their EHR. The higher rating score indicates a more appropriate composition. **b**, From a total of $n = 192$ comparisons from ten healthcare team members, TPN2.0 received the highest experts' rating scores. The scores are also significantly higher (Mann–Whitney U test, two-sided P value < 0.0001) than the actual prescribed TPN and random TPN. Data are presented as mean values ± s.e.m.

as levocarnitine, which is usually constrained to either 0, 10 or 20 mg kg^{-1} day^{-1}. Overall, this demonstrates that TPN2.0 performs similarly to experts in both institutions.

Blinded providers rate TPN2.0 higher than best practice

We validated TPN2.0 in a blinded study. We recruited ten members of the healthcare team to first chart review randomly selected NICU patients who received TPN. Next, they were presented with three potential TPN solutions: TPN2.0's recommendation, the actual TPN solution developed by the care team according to best practice and a random TPN prescription as a negative control. Their task was to rate each of the three options on a scale of 0 to 100, with 100 being the highest preference (Fig. 4a). A total of 192 comparisons were completed. The results showed that TPN2.0 received the highest average expert rating (score of 56), outperforming both prescribed (average score of 35) and random TPN (average score of 20; Fig. 4b). By counts, TPN2.0 was also rated the highest most frequently (Supplementary Fig. 3). To reduce possible individual human errors, we performed an additional analysis where we included only patients evaluated by at least three stakeholders and took the mean of their scores. Even when compared against this mixture of human experts setup, TPN2.0 consistently received the highest ratings, with the score for every patient either matching or exceeding that of the prescribed TPN, underscoring the robustness of TPN2.0 preference (Extended Data Fig. 5). This blinded study supplements our performance metrics from the retrospective studies by providing additional insights from real-world scenarios.

Deviation from TPN2.0 is associated with adverse outcomes

The process of prescribing TPN is one of the most frequently cited errors among high-alert medications in NICUs, potentially impacting patient outcomes[22,23]. Given this, we next asked whether retrospective data could shed light on the potential value of TPN2.0 in reducing adverse clinical outcomes. We selected a set of 16 critical neonatal morbidities, along with the congenital heart disease (CHD) status, to investigate whether the deviation from TPN2.0 in TPN prescribed before the diagnosis date is associated with an increase in these outcomes. The relationship between these outcomes is displayed in Fig. 5a, where closer nodes and thicker edges indicate the pairs that are more likely to co-occur. We calculated the odds ratio (OR) of each outcome when the actual prescriptions highly disagreed with TPN2.0 and displayed the results as a node size in Fig. 5a. Disagreement was defined as cases where the difference between the composition of TPN2.0 and the actual prescriptions fell within the top 20% of calculated distances across all pairs. Within this group of patients, the OR of mortality (fivefold), cholestasis (fivefold), NEC (threefold) and sepsis (twofold) increased significantly as the prescriptions deviated from TPN2.0 (Fig. 5b). These heightened ORs persisted even when adjusted for the volume of enteral feeds, if any, associated with the TPN prescriptions (Supplementary Fig. 4). Among these patients, actual prescriptions also exhibited higher variances among races than TPN2.0 (Extended Data Fig. 6). To avoid confounding factors, in each of these morbidities, we used control patients with matched gestational age, birth weight and postmenstrual age.

Over the decade of data collection, substantial changes in preterm infant care occurred. Key milestones include: 2013 (O_2 saturation goals changed from 88–92% to 90–95% (ref. 41) and CHD screening was introduced[42]), 2014 (cooling on transport for HIE[43]), 2015 (brain care bundle, small baby unit, weight-based postextubation support), 2016 (standardized intubation and premedication[44]), 2017 (Premiloc protocol, TPN cessation at 100 ml kg^{-1} feeds[45]) and 2018 (PDA closure with Piccolo[46,47]). Later changes, including less invasive surfactant administration/minimally invasive surfactant therapy and faster NEC feeding pathways, were implemented after the study period[48,49]. Notably, there have been no great changes in the way that TPN has been prescribed or administered over the study period. More importantly, these changes did not affect the conclusion of our study. Even if we stratified the cohort into three different periods, we still observe the increased ORs when prescriptions deviated from TPN2.0 for all four morbidities across all periods (Supplementary Fig. 5). This analysis supports that these changes in infant treatments during these periods had limited effect on the outcomes related to TPN2.0.

As opposed to disagreement with TPN2.0, we also analyzed cases with high similarity to TPN2.0 compared to other prescriptions. Similarity was defined as prescriptions where the calculated distance between TPN2.0 and the actual formulation fell within the bottom 20th percentile of distances across all pairs. TPN prescriptions that were highly aligned with TPN2.0 were associated with significantly lower ORs: mortality (fivefold reduction), cholestasis (twofold

reduction) and sepsis (twofold reduction; Supplementary Fig. 6). Collectively, these findings suggest that deviations from TPN2.0 are associated with increased odds of adverse outcomes, while close adherence could provide further benefits and reduce the likelihood of these complications.

In a separate analysis that combines these two insights, we used the same definition defining disagreement (>80th percentile) and similarity (<20th percentile) to TPN2.0 to conduct survival analysis of these outcomes. The analysis further emphasizes our previous results. For example, at the same distance proportion, the cholestasis survival rate of patients with prescribed TPN close to TPN2.0 was ~80% when those with prescribed TPN disagreeing with TPN2.0 were down to 50% (Fig. 5c).

An example patient is visualized in Fig. 5d. This patient showed a high deviation from TPN2.0 recommendations (more than the 95th percentile away) and was diagnosed with cholestasis, with fat dose accounting for one of the largest portions of this deviation before diagnosis (~2 g kg^{-1} in TPN2.0 as opposed to 3 g kg^{-1} in the actual prescriptions). A higher fat dose was probably prescribed to maximize growth and development. However, studies have also shown that excessive fat amounts are associated with heightened risks of cholestasis, especially in the prolonged use of soybean-based oil as in this case[50–53]. Note that, after the diagnosis, the actual prescription's fat doses were decreased to a level similar to that initially recommended by TPN2.0. This example illustrates one source of disagreement between TPN2.0 and actual prescriptions. Interestingly, an additional blinded study of patients with cholestasis revealed that deviations from TPN2.0 were unlikely to result from physicians intentionally modifying TPN compositions in anticipation of clinical outcomes. In the blinded test, the healthcare teams were unable to distinguish between TPN2.0 prescriptions and best-practice prescriptions before the diagnosis date, suggesting that any observed differences in outcomes were not confounded by preemptive changes in TPN formulations (Supplementary Fig. 7). This suggests that TPN compositions can be optimized to improve outcomes, providing a foundation for future studies to establish new gold standards in patient care. However, it is important to note that these inferences are based on observational data and are hypothesis-generating. While the best attempts were made to account for possible confounders, these findings should not be considered causal until confirmed by a randomized controlled study. TPN components that largely contribute to the overall difference between TPN2.0 and actual prescriptions are visualized in Supplementary Fig. 8, and the complex interplay between different nutrients and patient characteristics is visualized in Supplementary Fig. 9.

Enabling guideline-adhering and physician-in-the-loop TPN2.0

In real-world NICUs, interactions with physicians and strict adherence to safety guidelines are crucial. To address this, we next developed a transformer-based, physics-informed (PI) model that not only incorporates clinical pharmacy safety guidelines, but also enables physician-in-the-loop collaborative decision-making. This hybrid approach balances automation with human expertise, enhancing both safety and personalization in neonatal care.

When compounding TPN, the prescribed TPN dosage should adhere to clinical and institutional guidelines[54], such as osmolarity limits, maximum component concentrations and ensuring solubility. Indeed, the resulting TPN should also abide by physical expectations, such as the balance between total negative and positive ions. The list of all expectations and guidelines considered, as well as their limits, is tabulated in Supplementary Table 2. Here, we explored sequential models such as PI-transformers to achieve this goal. A transformer is a deep-learning model that uses an encoder to process longitudinal input data, that is, clinical characteristics, and a decoder that takes the previous days' ground truth for predictions, that is, TPN prescriptions, along with the encoded data to generate the current output (Fig. 6a). Adding to these structures, the PI-transformer was designed to adhere to specific constraints while making predictions. Figure 6b presents the average violation across all criteria, with individual criteria detailed in Extended Data Fig. 7. The PI-transformer exhibited ~sixfold lower violation of these rules compared to the normal transformer architecture, and was the lowest among all algorithms considered. It also achieved this without sacrificing general correlation performance compared to the normal transformer or any other cutting-edge deep-learning methods (Extended Data Fig. 8).

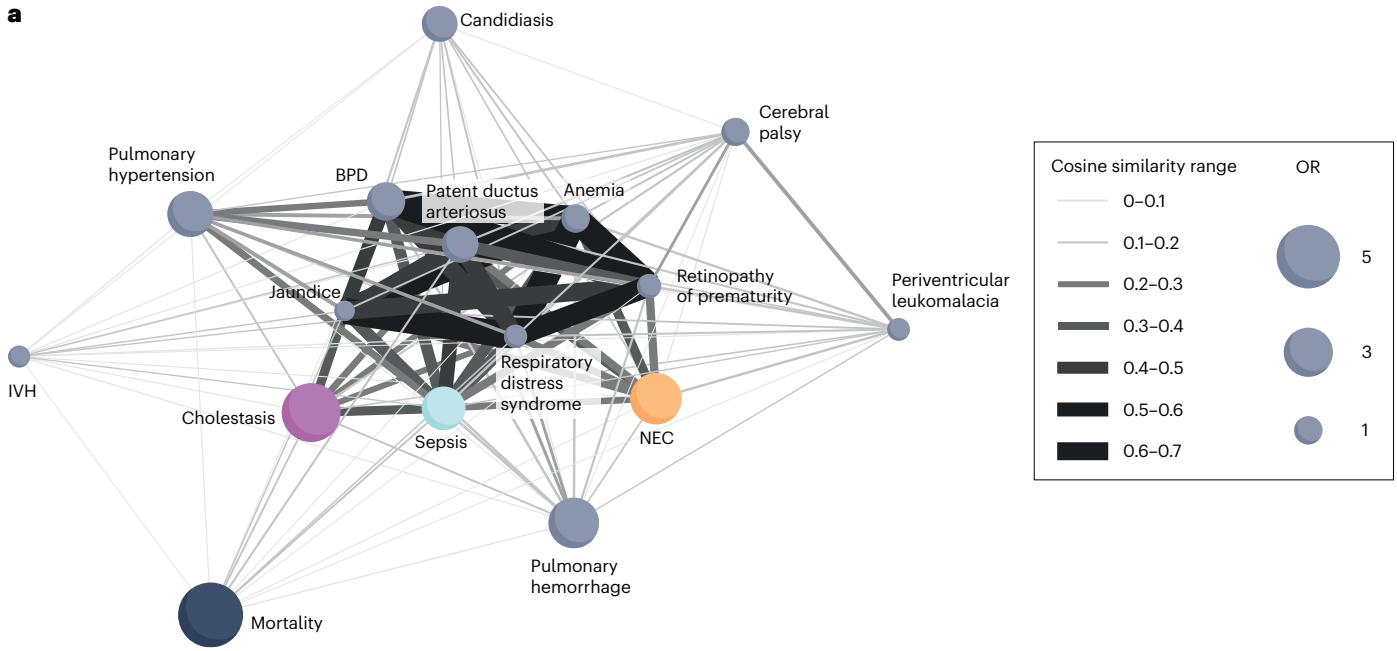
In clinical settings, exceptional cases are inevitable, making it essential that future versions of TPN2.0 are flexible enough to work collaboratively with physicians to respond to unique patient needs. To address this, we aim to explore the PI-transformer's ability to incorporate physicians' demands. We implemented a custom training method by using teacher forcing only for pretraining, followed by inference as training to fine-tune the model. This resulted in a substantial improvement to a correlation of 0.66. This TPN-specific training method not only improved performance but also enabled the transformer to meet physicians' demands. For comparison, the baseline performance using teacher forcing is only 0.48 (Fig. 6c). The model was also interpretable end-to-end, allowing physicians to identify features that drive the model predictions both at the global and local (individual) levels. It also highlights how a single feature can influence predictions differently across samples. For example, in some cases, higher serum calcium increases the likelihood of being assigned to a specific cluster. However, in other cases, it has the opposite effect (Extended Data Fig. 9).

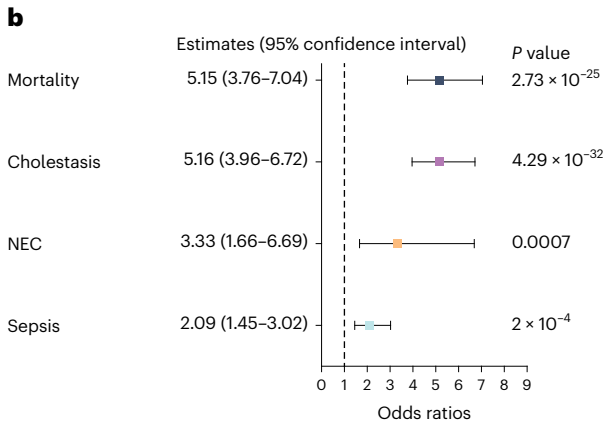
To highlight the utility of this approach, consider a scenario where a physician disagrees with components of the recommendation from the transformer. The physician could modify the recommendation. The next day, the model can use this modified prescription as the decoder's input instead of its previous day's prediction.

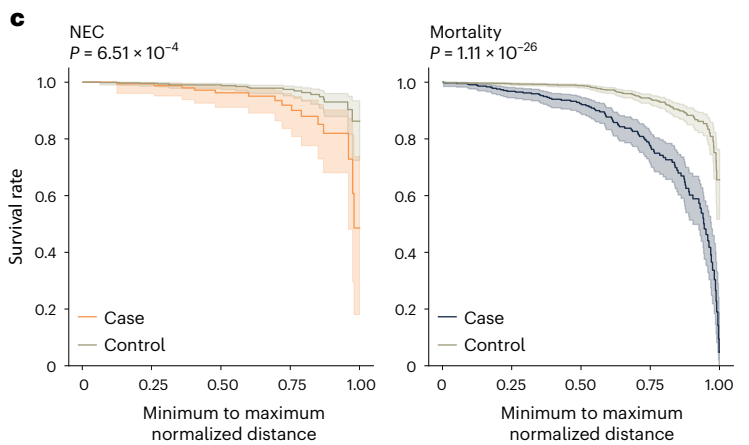
Fig. 5 | TPN2.0 recommendations are associated with lower rates of morbidities and mortality. a, A correlation network visualizing the cosine similarity between 16 neonatal outcomes. The size of each node is proportional to the OR of developing the morbidity when the patient's prescriptions deviated from TPN2.0. Prescriptions are considered to deviate from TPN2.0 when the average Manhattan distance between their compositions is more than the 80th percentile away. Only prescriptions before the diagnosis (up to 3 months) are included. The thickness and color of the edges are proportional to the strength of the cosine similarity; thicker and darker lines indicate higher similarity. **b**, ORs in patients who received prescriptions deviating from TPN2.0. Morbidities that showed a significant increase in OR include cholestasis, NEC, sepsis and mortality. The *P* value of the OR is obtained through a two-sided *z* test for the log OR and is not adjusted for multiple comparison. Dots, mean values; error bars, 95% confidence intervals. **c**, Survival plots depicting the difference in the rate of developing an outcome between cases and controls as the distance between TPN2.0 and the actual prescription grows. The case group consists of patients whose average distances between TPN2.0 and the actual prescriptions are beyond the 80th percentile, that is, those with prescribed TPN with composition very different from TPN2.0. The control group consists of those with distances below the 20th percentile, that is, those with prescribed TPN with very similar compositions to TPN2.0 recommendations. Lines, estimated survival probability; shading, 95% confidence intervals. The *P* value is obtained from log rank test. **d**, To demonstrate this with an example, a patient whose actual prescriptions deviated from TPN2.0 and developed cholestasis on day 64 of TPN is visualized. The difference in fat dosage is one of the main contributors to the deviation from TPN2.0 recommendations. The actual prescriptions have mostly 3 g kg^{-1} while TPN2.0 recommends 2 g kg^{-1} before the diagnosis. In past studies, restriction of TPN fat in vulnerable populations has been associated with reduced risks of cholestasis[50–53]. DBili, direct bilirubin.

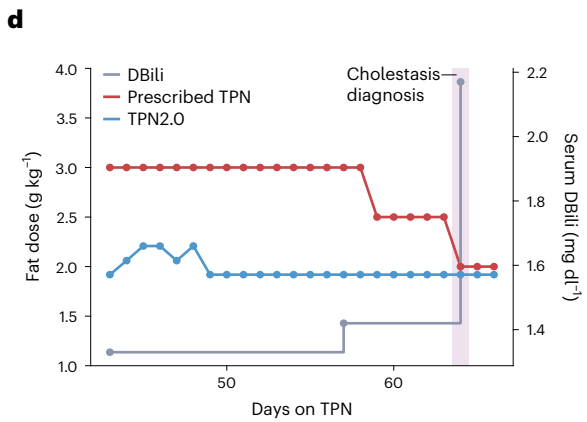
https://doi.org/10.1038/s41591-025-03601-1

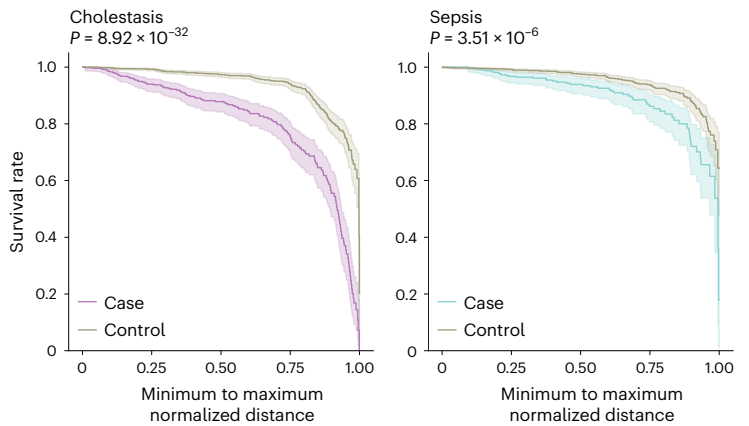
This modification would allow the model to better mimic a given prescriber. To evaluate the model, we replicated this scenario by replacing TPN2.0 compositions with actual prescriptions in our dataset. By replacing only 5% of the TPN2.0 recommendations (with the biggest disagreement with physicians) with actual prescriptions, the correlation between TPN2.0 recommendations and prescribed TPN on the next days increased to 0.72 (Fig. 6c). The trend continued with the frequency of physician intervention and reached a correlation of 0.78 at 20% intervention. Further analyses showed that this improvement was observed across all individual physicians (Extended Data Fig. 10) and that the model still preserved most of its associations to improved outcomes even after intervention (Supplementary Fig. 10). This demonstrated how AI can synergistically co-pilot with healthcare providers providing real-world NICU care.











https://doi.org/10.1038/s41591-025-03601-1

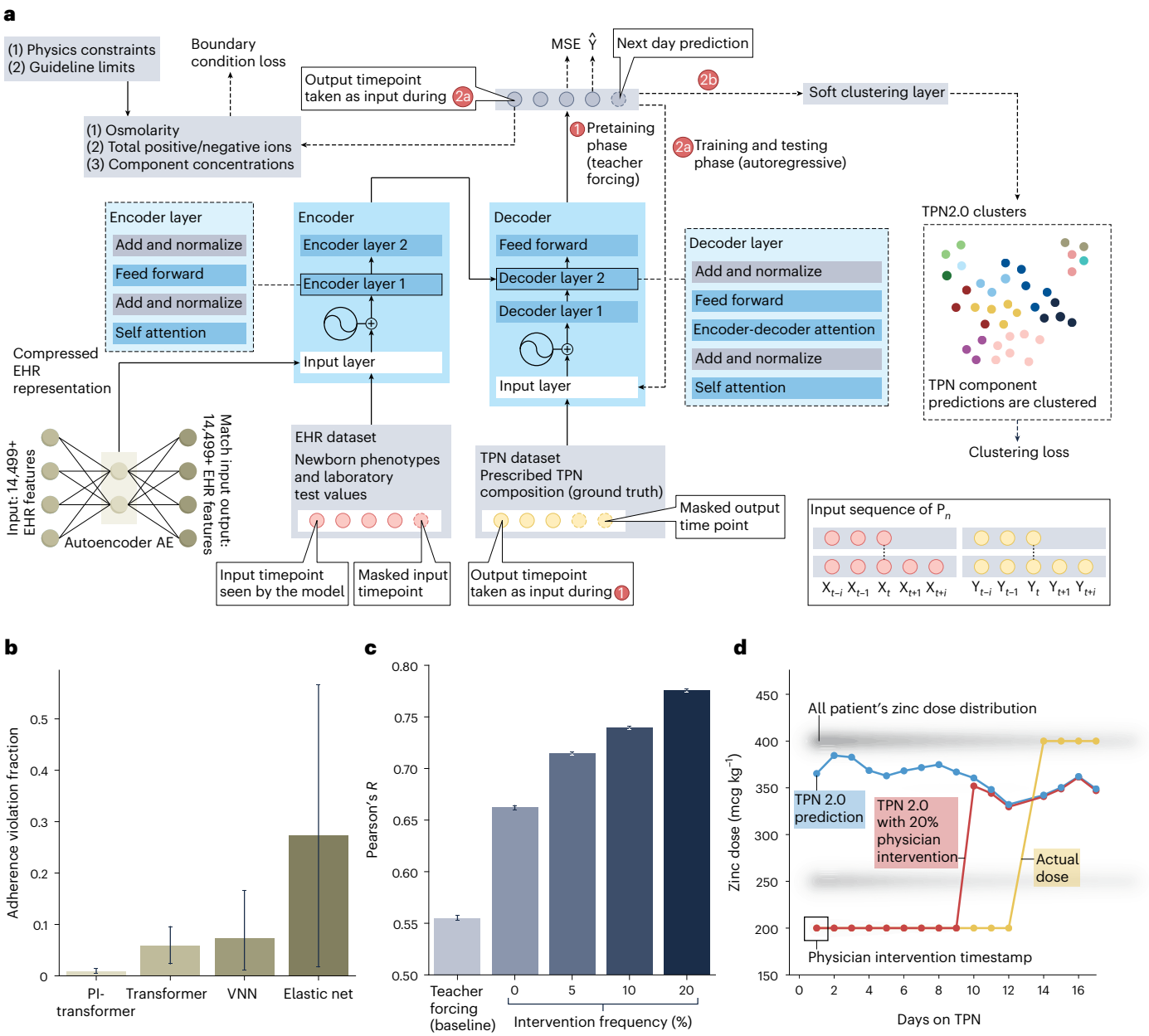


Fig. 6 | TPN2.0 by PI-transformer enables physician-in-the-loop recommendations that adhere to pharmacist guidelines. a, A PI-transformer was developed to cluster TPN compositions over time. To predict TPN composition at time t ($\hat{Y}_t$) for patient P_n, future data (X_{t+i}) is masked to prevent information leakage. The model employs a positional encoder for daily TPN to generate latent representations that the decoder combines with previous TPN data (Y_{t-i}) to predict $\hat{Y}_t$. In pretraining, teacher forcing is used; in fine-tuning, 'inference as training' is applied as the decoder autoregressively processes previous predictions. During inference, the model predictions could also be replaced by actual prescriptions if needed. The predictions are further utilized to calculate TPN characteristics. **b**, TNP2.0 recommendations comply with pharmacist guidelines and rules. These computed values, together with ten pharmacist guidelines/physical expectations (Supplementary Table 2)—including osmolarity, dextrose concentrations and calcium phosphate solubility limits—are integrated into boundary condition losses to enforce clinical standards. TPN2.0 with PI-transformer exhibited the fewest violations among all algorithms tested (n = 79,790 prescriptions from 5,913 patients). **c**, The performance of TPN2.0 improves with increased physician intervention. Simulated interventions, in which 10% of the TPN2.0 recommendations that are least consistent with actual prescriptions are replaced by the actual prescriptions' values in the decoder, further enhance performance. At 0% intervention, the model also outperforms the baseline teacher forcing method. This analysis mimics real-world scenarios where physicians modify AI recommendations, and shows that closer collaboration between AI and clinicians enhances model accuracy. Data are presented as mean values ± s.e.m. **d**, In one illustrative case, the model's zinc prediction of 362 mcg kg^{-1} on day 1 was modified to 200 mcg kg^{-1} according to the actual prescription. The gray area represents the distribution of all zinc values in the data. After the intervention, the model maintained a zinc of 200 mcg kg–1 for the following 8 days, consistent with the actual prescriptions. Subsequently, the prediction shifted back to approximately 350 mcg kg^{-1}—a change that was later followed by the physicians. This indicates the ability of the model to balance clinical judgment with the data-driven approach.

An example of this intervention is shown in Fig. 6d where we simulated the physician modifying a recommendation due to the large distance between TPN2.0 and the actual prescription. Zinc was the main difference between the recommendation and the physician's preference. Here, the transformer suggested a zinc dosage of 364 mcg kg^{-1}, whereas the actual prescription was 200 mcg kg^{-1}. After the simulated intervention, the transformer was able to adapt and kept zinc at the same level the physician would have wanted for

the next 8 days. According to guidelines, preterm infants like this case should receive 400 mcg kg^{-1} of zinc[55]. TPN2.0 correctly recommended the higher zinc, but in a rare circumstance such as this, providers may prescribe less zinc to accommodate other unique clinical factors. Over time, the transformer shifted zinc back to a higher level, which the actual prescriptions then followed only 3 days later. This result showcased that TPN2.0 can adapt to a provider's needs while balancing consistency with its data-driven recommendations.

Discussion

We aggregated a cohort of 79,790 TPN orders linked to EHRs from 5,913 neonates and an external cohort with 63,273 TPN orders from 3,417 newborns and infants. Our data, collected over 10 years, include TPN orders with a nearly equal distribution of sexes and several racial representations (Supplementary Table 1). From these comprehensive datasets, we demonstrated that an AI model can develop TPN compositions in a data-driven manner. Based on the model, we also report a set of AI-driven TPN formulas that could respond to the varying nutritional needs of newborns while still being manageable for mass production, substantially reducing cost and risk for errors. Although the racial distribution modestly underrepresented Black/African American and Indigenous populations, findings from stratified analyses suggested our results were broadly applicable across racial groups (Extended Data Fig. 6). We then showed that TPN2.0 recommendations were rated higher by experts compared to current best practice, and validated the results independently in a second hospital system. Through in-depth analyses of newborn outcomes, we identified associations between the onset of morbidities and deviation from TPN2.0. We also leveraged a PI-transformer architecture to make our model adaptive to input from physicians and adhere to clinical/pharmacy safety guidelines. These benefits allowed TPN2.0 to standardize TPN with a personalization level comparable with the state-of-the-art, while also potentially making high-quality TPN potentially more accessible, both in low-resource settings in the United States and in LMICs.

While numerous predictive models have been deployed in intensive care settings[56,57], often with great accuracy, these predictions have yet to meaningfully change clinical outcomes or practices[12]. In the current healthcare environment where the care team has to balance patient care with information overload and alarm fatigue, AI models need to focus on actionability, not just predictive modeling[58,59]. TPN presents a unique opportunity for such actionability, in partnership between AI and physicians. Currently, administering TPN for a neonate is challenging due to their complex nutritional needs and lack of clear clinical guidelines[60], resulting in overreliance on anecdotal experience. Current TPN practice also requires an extensive amount of time and human resources to prescribe and prepare. With many stakeholders involved, TPN is error-prone, with 65% of errors occurring even before administration to patients[61]. The data-driven approach, standardization and objective recommendations provided by TPN2.0 directly address these limitations.

The benefits of standardized TPN, such as reduction in error, cost and practice variation[62], are well established. Indeed, the industry, as well as professional societies worldwide[20,63–65], have been attempting to move in this direction. Standardized TPN solutions are typically developed based on expert consensus and are considered best practice in some countries, such as the UK and Australia[63,64,66]. However, the lack of rigorous evidence through randomized trials and their design by expert opinion as a 'gold standard' raises concerns about their applicability, particularly for preterm infants—a population with highly diverse and individualized nutritional needs. Reflecting this, organizations like ASPEN advised against the use of standardized TPN in such cases[20]. The other bottleneck toward this goal is the lack of personalization in which the current one(or few)-size-fits-all solutions could jeopardize patient safety, such as by causing electrolyte disorders[67–69] including hyponatremia or hypermagnesemia[70]. This drawback has prevented the widespread adoption of standardized TPN in many countries, including the United States. In contrast, customized TPN based on an individual expert opinion may introduce biases stemming from anecdotal experience. TPN2.0 strikes a balance between personalization and standardization by leveraging a data-driven approach, combining the strengths of both methods. Its recommendations were grounded in evidence and were associated with improved outcomes, such as reduced morbidity rates. The 15 standardized formulas produced by TPN2.0 increase safety and reduce cost, while the AI model maintains a high degree of responsiveness to patients' dynamic needs during NICU stay.

TPN2.0 was designed to collaborate with the healthcare team in clinical settings to promote ease of adoption and increase safety. AI algorithms focused purely on a black box approach to clinical predictions often struggle with adoption and generalizability in the real world. In contrast, models with physician-in-the-loop have been evaluated to have improved accuracy and user satisfaction[71,72]. TPN2.0 has been developed to adapt to individual physicians' judgment when needed, while continuing to promote a data-driven approach. This ensures that the AI serves as a supportive tool for physicians, while leaving the ultimate decision-making authority firmly in their hands. TPN2.0 was also designed to automatically adhere to clinical/pharmacy guidelines and composition requirements. Although validations in clinical settings are still required before the deployment of TPN2.0, these initial findings already demonstrate its ability to cater to both patient nutritional needs and clinical requirements.

One limitation of this study is the reliance on structured EHR data, which lacks comprehensive patient information and is prone to dataset shift[73]. This shift emphasizes the need for continuous AI performance monitoring, as highlighted by the United States Food and Drug Administration[74]. Although our dataset includes key TPN-related variables (laboratory values, diagnoses and procedures), it excludes imaging, waveforms and clinical notes. Future studies should incorporate multimodal or biological systems for enhanced decision-making[75,76]. Our blinded study was limited to Stanford experts prescribing TPN. To validate TPN2.0 preferences, future research will involve experts from other institutions to explore the impact of training, experience and local protocols. Variations in enteral feed types (formula versus breast milk) and compositions remain unexplored and should be investigated in future studies. Ethical considerations are critical when applying AI in the NICU, as decision-making involves parents or guardians representing critically ill newborns. Transparency and alignment with clinical expertise are essential to support, not replace, decision-making. Real-world data biases may also confound findings, as specific TPN compositions could be prescribed for particular conditions and outcomes. Although we mitigated potential confounders by analyzing prediagnosis TPN data and using a blinded study the findings remain preliminary and associative and need prospective validation in a randomized controlled trial (RCT). RCTs are crucial to establish causality, control biases and provide stronger clinical evidence. Future RCTs should assess not only growth and development but also nutrition-related morbidities (for example, hyponatremia, hypermagnesemia) and long-term outcomes such as neurodevelopment[77] and the incidence of chronic conditions like metabolic and cardiovascular disorders[78,79]. Despite current limitations, this study highlights the ground-breaking potential of TPN2.0 to improve neonatal health, offering life-long benefits that are often harder to achieve in older populations.

Taken together, TPN2.0 demonstrates the potential of AI to go beyond predictive diagnosis and guide a key therapeutic decision for newborns in the most vulnerable time of their lives. This approach streamlines and standardizes TPN processes, enhancing accessibility (particularly in low-resource settings) and increasing safety. Our work sets the foundation for future studies to assess the real-world impact of TPN2.0 on this patient population, which may be influenced

by various confounding factors. These studies will include ongoing efforts for multisite model validation and prospective clinical trials. This framework also allows future incorporation of causal AI[80,81]. This can better inform the design and execution of RCTs to establish a true gold standard for TPN that is based objectively on the improved outcomes for individual newborn characteristics. It also highlights the fundamental role that AI will have in future neonatal intensive care medicine. This progress will lay the foundation for AI models to autonomously guide therapeutic strategies beyond TPN, paving the way for the development of large action models[82].

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41591-025-03601-1.

References

1. Granger, C. L., Okpapi, A., Peters, C. & Campbell, M. G578(P) Potentially preventable unexpected term admissions to neonatal intensive care (NICU). *Arch. Dis. Child.* **100**, A263–A263 (2015).
2. Harrison, W. & Goodman, D. Epidemiologic trends in neonatal intensive care, 2007–2012. *JAMA Pediatr.* **169**, 855–862 (2015).
3. Pang, E. M. et al. Evaluating epidemiologic trends and variations in NICU admissions in California, 2008 to 2018. *Hosp. Pediatr.* **13**, 976–983 (2023).
4. Martin, J. A. & Osterman, M. J. K. Shifts in the distribution of births by gestational age: United States, 2014–2022. *Natl Vital Stat. Rep.* **73**, 1–11 (2024).
5. Gregory, G. A., Kitterman, J. A., Phibbs, R. H., Tooley, W. H. & Hamilton, W. K. Treatment of the idiopathic respiratory-distress syndrome with continuous positive airway pressure. *N. Engl. J. Med.* **284**, 1333–1340 (1971).
6. Gitlin, J. D. et al. Randomized controlled trial of exogenous surfactant for the treatment of hyaline membrane disease. *Pediatrics* **79**, 31–37 (1987).
7. Meeus, M. et al. Clinical decision support for improved neonatal care: the development of a machine learning model for the prediction of late-onset sepsis and necrotizing enterocolitis. *J. Pediatr.* **266**, 113869 (2024).
8. Yang, Y.-H. et al. Predicting early mortality and severe intraventricular hemorrhage in very-low birth weight preterm infants: a nationwide, multicenter study using machine learning. *Sci. Rep.* **14**, 10833 (2024).
9. Wu, K. et al. Characterizing the clinical adoption of medical AI devices through U.S. insurance claims. *NEJM AI* **1.1**, AIoa2300030 (2024).
10. Artificial intelligence and machine learning (AI/ML)-enabled medical devices. *United States Food and Drug Administration* www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices (2024).
11. Lee, S. E., Hong, H. & Kim, E.-K. Diagnostic performance with and without artificial intelligence assistance in real-world screening mammography. *Eur. J. Radiol. Open* **12**, 100545 (2024).
12. Ahmad, T. et al. Alerting clinicians to 1-year mortality risk in patients hospitalized with heart failure: the REVEAL-HF randomized clinical trial. *JAMA Cardiol.* **7**, 905–912 (2022).
13. Huang, Q. et al. Position paper: agent AI towards a holistic intelligence. Preprint at https://arxiv.org/abs/2403.00833v1 (2024).
14. Brennan, M. R. The wild west of AI-powered devices. *Device* **2**, 100364 (2024).
15. Chaudhari, S. & Kadam, S. Total parenteral nutrition in neonates. *Indian Pediatr.* **43**, 953–964 (2006).
16. Skouroliakou, M. et al. Comparison of two types of TPN prescription methods in preterm neonates. *Pharm. World Sci.* **31**, 202–208 (2009).
17. Turpin, R. S., Liu, F. X., Prinz, M., Macahilig, C. & Malinoski, F. Parenteral nutrition prescribing pattern. *Nutr. Clin. Pract.* **28**, 242–246 (2013).
18. Harding, J. E., Cormack, B. E., Alexander, T., Alsweiler, J. M. & Bloomfield, F. H. Advances in nutrition of the newborn infant. *Lancet* **389**, 1660–1668 (2017).
19. ElHassan, N. O. & Kaiser, J. R. Parenteral nutrition in the neonatal intensive care unit. *Neoreviews* **12**, e130–e140 (2011).
20. Robinson, D. T. et al. Guidelines for parenteral nutrition in preterm infants: the American Society for Parenteral and Enteral Nutrition. *J. Parenter. Enteral Nutr.* **47**, 830–858 (2023).
21. Mihatsch, W. et al. ESPGHAN/ESPEN/ESPR/CSPEN guidelines on pediatric parenteral nutrition: guideline development process for the updated guidelines. *Clin. Nutr.* **37**, 2306–2308 (2018).
22. Stavroudis, T. A. et al. NICU medication errors: identifying a risk profile for medication errors in the neonatal intensive care unit. *J. Perinatol.* **30**, 459–468 (2010).
23. Aschenbrenner, D. S. ISMP updates high-alert medication list for acute care settings. *Am. J. Nurs.* **124**, 20 (2024).
24. Lehmann, C. U., Conner, K. G. & Cox, J. M. Preventing provider errors: online total parenteral nutrition calculator. *Pediatrics* **113**, 748–753 (2004).
25. Stultz, J. S. & Nahata, M. C. Computerized clinical decision support for medication prescribing and utilization in pediatrics. *J. Am. Med. Inform. Assoc.* **19**, 942–953 (2012).
26. Abacus Software. *Baxter* https://ushospitalproducts.baxter.com/abacus-software (2020).
27. Peverini, R. L., Beach, D. S., Wan, K. W. & Vyhmeister, N. R. Graphical user interface for a neonatal parenteral nutrition decision support system. *Proc. AMIA Symp.* **2000**, 650–654 (2000).
28. Monterey Medical Solutions, Inc. *Medical Device Network* http://lucee.warrick.net/companies/index.cfm?fuseaction=companies_detail&eregnum=3004967992
29. Lashinsky, J. N., Suhajda, J. K., Pleva, M. R. & Kraft, M. D. Use of integrated clinical decision support tools to manage parenteral nutrition ordering: experience from an academic medical center. *Nutr. Clin. Pract.* **36**, 418–426 (2021).
30. Webbe, J. W. H. et al. Outcomes in relation to early parenteral nutrition use in preterm neonates born between 30 and 33 weeks' gestation: a propensity score matched observational study. *Arch. Dis. Child. Fetal Neonatal. Ed.* **107**, 131–136 (2022).
31. Uthaya, S. et al. Early versus later initiation of parenteral nutrition for very preterm infants: a propensity score-matched observational study. *Arch. Dis. Child. Fetal Neonatal Ed.* **107**, 137–142 (2022).
32. Moon, K. & Rao, S. C. Early or delayed parenteral nutrition for infants: what evidence is available? *Curr. Opin. Clin. Nutr. Metab. Care* **24**, 281–286 (2021).
33. Valentine, C. J. et al. Early amino-acid administration improves preterm infant weight. *J. Perinatol.* **29**, 428–432 (2009).
34. Moon, K., Athalye-Jape, G. K., Rao, U. & Rao, S. C. Early versus late parenteral nutrition for critically ill term and late preterm infants. *Cochrane Database Syst. Rev.* **4**, CD013141 (2020).
35. Battersby, C. et al. Incidence and enteral feed antecedents of severe neonatal necrotising enterocolitis across neonatal networks in England, 2012-13: a whole-population surveillance study. *Lancet Gastroenterol. Hepatol.* **2**, 43–51 (2017).
36. Huang, J. et al. Human milk as a protective factor for bronchopulmonary dysplasia: a systematic review and meta-analysis. *Arch. Dis. Child. Fetal Neonatal Ed.* **104**, F128–F136 (2019).

37. *Guidelines on optimal feeding of low-birth-weight infants in low- and middle-income countries* (World Health Organization, 2011); www.who.int/publications/i/item/9789241548366
38. Gidi, N. W. et al. Preterm nutrition and clinical outcomes. *Glob. Pediatr. Health* **24**, 2333794X20937851 (2020).
39. ESPGHAN and ESPEN guidelines paediatric parenteral nutrition—annex: list of products. *J. Pediatr. Gastroenterol. Nutr*. **41**, S85–S87 (2005).
40. Senterre, T. et al. Safe and efficient practice of parenteral nutrition in neonates and children aged 0–18 years—the role of licensed multi-chamber bags. *Clin. Nutr.* **43**, 1696–1705 (2024).
41. SUPPORT Study Group of the Eunice Kennedy Shriver NICHD Neonatal Research Network et al.Target Ranges of Oxygen Saturation in Extremely Preterm Infants. *N. Engl. J. Med.* **362**, 1959–1969 (2010).
42. Kemper, A. R. et al. Strategies for implementing screening for critical congenital heart disease. *Pediatrics* **128**, e1259–e1267 (2011).
43. Fairchild, K., Sokora, D., Scott, J. & Zanelli, S. Therapeutic hypothermia on neonatal transport: 4-year experience in a single NICU. *J. Perinatol.* **30**, 324–329 (2009).
44. Shay, R., Weikel, B. W., Grover, T. & Barry, J. S. Standardizing premedication for non-emergent neonatal tracheal intubations improves compliance and patient outcomes. *J. Perinatol.* **42**, 132–138 (2021).
45. Baud, O. et al. Effect of early low-dose hydrocortisone on survival without bronchopulmonary dysplasia in extremely preterm infants (PREMILOC): a double-blind, placebo-controlled, multicentre, randomised trial. *Lancet* **387**, 1827–1836 (2016).
46. Zahn, E. M., Nevin, P., Simmons, C. & Garg, R. A novel technique for transcatheter patent ductus arteriosus closure in extremely preterm infants using commercially available technology. *Catheter. Cardiovasc. Interv.* **85**, 240–248 (2015).
47. Sathanandam, S. K. et al. Amplatzer Piccolo Occluder clinical trial for percutaneous closure of the patent ductus arteriosus in patients ≥700 grams. *Catheter. Cardiovasc. Interv.* **96**, 1266–1276 (2020).
48. Herting, E. Less invasive surfactant administration (LISA)—ways to deliver surfactant in spontaneously breathing infants. *Early Hum. Dev.* **89**, 875–880 (2013).
49. Herting, E., Härtel, C. & Göpel, W. Less invasive surfactant administration (LISA): chances and limitations. *Arch. Dis. Child. Fetal Neonatal Ed.* **104**, F655–F659 (2019).
50. Guthrie, G. & Burrin, D. Impact of Parenteral Lipid Emulsion Components on Cholestatic Liver Disease in Neonates. *Nutrients* **13**, 508 (2021).
51. Cober, M. P. et al. Intravenous fat emulsions reduction for patients with parenteral nutrition-associated liver disease. *J. Pediatr.* **160**, 421–427 (2012).
52. Sanchez, S. E. et al. The effect of lipid restriction on the prevention of parenteral nutrition-associated cholestasis in surgical infants. *J. Pediatr. Surg.* **48**, 573–578 (2013).
53. Rollins, M. D. et al. Effect of decreased parenteral soybean lipid emulsion on hepatic function in infants at risk for parenteral nutrition-associated liver disease: a pilot study. *J. Pediatr. Surg.* **48**, 1348–1356 (2013).
54. Boullata, J. I. et al. A.S.P.E.N. clinical guidelines. *JPEN J. Parenter. Enteral Nutr.* **38**, 334–377 (2014).
55. Friel, J. K. & Andrews, W. L. Zinc requirement of premature infants. *Nutrition* **10**, 63–65 (1994).
56. Crilly, C. J., Haneuse, S. & Litt, J. S. Predicting the outcomes of preterm neonates beyond the neonatal intensive care unit: what are we missing? *Pediatr. Res.* **89**, 426–445 (2021).
57. Beam, K., Sharma, P., Levy, P. & Beam, A. L. Artificial intelligence in the neonatal intensive care unit: the time is now. *J. Perinatol.* **44**, 131–135 (2024).
58. Sbaffi, L., Walton, J., Blenkinsopp, J. & Walton, G. Information overload in emergency medicine physicians: a multisite case study exploring the causes, impact, and solutions in four North England National Health Service Trusts. *J. Med. Internet Res.* **22**, e19126 (2020).
59. Li, C., Parpia, C., Sriharan, A. & Keefe, D. T. Electronic medical record-related burnout in healthcare providers: a scoping review of outcomes and interventions. *BMJ Open* **12**, e060865 (2022).
60. Kuzma-O'Reilly, B. et al. Evaluation, development, and implementation of potentially better practices in neonatal intensive care nutrition. *Pediatrics* **111**, e461–e470 (2003).
61. Sacks, G. S. Safety surrounding parenteral nutrition systems. *JPEN J. Parenter. Enteral Nutr.* **36**, 20S–22S (2012).
62. Rigo, J. et al. Benefits of a new pediatric triple-chamber bag for parenteral nutrition in preterm infants. *J. Pediatr. Gastroenterol. Nutr.* **54**, 210–217 (2012).
63. Bolisetty, S., Osborn, D., Sinn, J., Lui, K. & Australasian Neonatal Parenteral Nutrition Consensus Group. Standardised neonatal parenteral nutrition formulations—an Australasian group consensus 2012. *BMC Pediatr.* **14**, 48 (2014).
64. Riskin, A., Picaud, J.-C., Shamir, R. & ESPGHAN/ESPEN/ESPR/CSPEN working group on pediatric parenteral nutrition. ESPGHAN/ESPEN/ESPR/CSPEN guidelines on pediatric parenteral nutrition: standard versus individualized parenteral nutrition. *Clin. Nutr.* **37**, 2409–2417 (2018).
65. *Standardised neonatal parenteral nutrition formulations*. NICE Guideline No. 154 (National Institute for Health and Care Excellence, 2020); www.ncbi.nlm.nih.gov/books/NBK555683/
66. Bolisetty, S. et al. Standardised neonatal parenteral nutrition formulations—Australasian neonatal parenteral nutrition consensus update 2017. *BMC Pediatrics* **20**, 59 (2020).
67. Hakeam, H., Alsemari, M., Mohamed, G., Alshahrani, A. & Islami, M. The rate of discontinuing ready-to-use multi-chamber bag parenteral nutrition secondary to high serum electrolyte levels. *Hosp. Pharm.* **58**, 263–271 (2023).
68. Garner, S. S. et al. The impact of 2 weight-based standard parenteral nutrition formulations compared with one standard formulation on the incidence of hyperglycemia and hypernatremia in low birth-weight preterm infants. *Adv. Neonatal Care* **21**, E65–E72 (2021).
69. Mihatsch, W. et al. Systematic review on individualized versus standardized parenteral nutrition in preterm infants. *Nutrients* **15**, 1224 (2023).
70. Arnell, H. et al. Safety of a triple-chamber bag parenteral nutrition in children ages up to 24 months: an observational study. *J. Pediatr. Gastroenterol. Nutr.* **69**, e151–e157 (2019).
71. Topol, E. J. High-performance medicine: the convergence of human and artificial intelligence. *Nat. Med.* **25**, 44–56 (2019).
72. Rajpurkar, P., Chen, E., Banerjee, O. & Topol, E. J. AI in health and medicine. *Nat. Med.* **28**, 31–38 (2022).
73. Finlayson, S. G. et al. The clinician and dataset shift in artificial intelligence. *N. Engl. J. Med.* **385**, 283–286 (2021).
74. Warraich, H. J., Tazbaz, T. & Califf, R. M. FDA perspective on the regulation of artificial intelligence in health care and biomedicine. *JAMA* **333**, 241–247 (2024).
75. Mataraso, S. J. et al. A machine learning approach to leveraging electronic health records for enhanced omics analysis. *Nat. Mach. Intell.* **7**, 293–306 (2025).
76. Seong, D. et al. Generating pregnant patient biological profiles by deconvoluting clinical records with electronic health record foundation models. *Brief. Bioinform.* **25**, bbae574 (2024).
77. Boscarino, G. et al. Neonatal hyperglycemia related to parenteral nutrition affects long-term neurodevelopment in preterm newborn: a prospective cohort study. *Nutrients* **13**, 1930 (2021).

78. Randunu, R. et al. Intrauterine growth-restricted piglets are predisposed to develop metabolic disorders in. adulthood when fed with parenteral nutrition in the neonatal period. *Curr. Dev. Nutr.* **6**, 703 (2022).
79. Elefson, S. K. et al. Adverse metabolic phenotypes in parenterally fed neonatal pigs do not persist into adolescence. *J. Nutr.* **154**, 638–647 (2024).
80. Feuerriegel, S. et al. Causal machine learning for predicting treatment outcomes. *Nat. Med.* **30**, 958–968 (2024).
81. Plecko, D. & Bareinboim, E. Causal fairness for outcome control. *Adv. Neural Inf. Process. Syst.* **36**, 47575–47597 (2023).
82. Zhang, J. et al. XLAM: A family of large action models to empower AI agent systems. Preprint at https://arxiv.org/abs/2409.03215 (2024).

Thanaphong Phongpreecha[1,2,3,4,11], Marc Ghanem[1,2,3,11], Jonathan D. Reiss[2,11], Tomiko T. Oskotsky[5,6,11], Samson J. Mataraso[1,2,3], Davide De Francesco[1,2,3], S. Momsen Reincke[1,2,3], Camilo Espinosa[1,2,3], Philip Chung[1], Taryn Ng[7], Jean M. Costello[5], Jennifer A. Sequoia[2], Sheila Razdan[2,8], Feng Xie[1,2,3], Eloise Berson[1,3,4], Yeasul Kim[1,2,3], David Seong[1,2,3], May Y. Szeto[2], Faith Myers[2], Hannah Gu[2], John Feister[2], Courtney P. Verscaj[2], Laura A. Rose[2], Lucas W. Y. Sin[1], Boris Oskotsky[5], Jacquelyn Roger[5], Chi-hung Shu[1,2,3], Sayane Shome[1,2,3], Liu K. Yang[1,2,3], Yuqi Tan[4,9], Steven Levitte[10], Ronald J. Wong[2], Brice Gaudillière[1], Martin S. Angst[1], Thomas J. Montine[4], John A. Kerner[2], Roberta L. Keller[6], Gary M. Shaw[2], Karl G. Sylvester[2], Janene Fuerch[2], Valerie Chock[2], Shabnam Gaskari[7], David K. Stevenson[2], Marina Sirota[5,6], Lawrence S. Prince[2] & Nima Aghaeepour[1,2,3] ✉

[1]Department of Anesthesiology, Pain and Perioperative Medicine, Stanford University, Stanford, CA, USA. [2]Department of Pediatrics, Stanford University, Stanford, CA, USA. [3]Department of Biomedical Data Science, Stanford University, Stanford, CA, USA. [4]Department of Pathology, Stanford University, Stanford, CA, USA. [5]Bakar Computational Health Sciences Institute, University of California, San Francisco, CA, USA. [6]Department of Pediatrics, University of California, San Francisco, CA, USA. [7]Department of Pharmacy, Lucile Packard Children's Hospital, Palo Alto, CA, USA. [8]Department of Pediatrics, Keck School of Medicine of USC, Division of Neonatal and Infant Critical Care Unit, Children's Hospital Los Angeles, Los Angeles, CA, USA. [9]Department of Microbiology and Immunology, Stanford University, Stanford, CA, USA. [10]Division of Pediatric Gastroenterology, Hepatology and Nutrition, Stanford University, Stanford, CA, USA. [11]These authors contributed equally: Thanaphong Phongpreecha, Marc Ghanem, Jonathan D. Reiss, Tomiko T. Oskotsky. ✉e-mail: naghaeep@stanford.edu

Methods

Study design

The discovery cohort data were aggregated from the EHRs at Stanford Health Care. The linkage of the two datasets enabled the integration of nutritional data, phenotypic traits and long-term outcomes. All Stanford EHR data were mapped to the Observational Medical Outcomes Partnership Common Data Model v.5.3.1. The data included patient observations, procedures, medications and conditions. The study was approved by the Institutional Review Board of Stanford University (reference no. 39225) with inform consent waived. We first identified a cohort of 6,991 neonatal/pediatric patients in NICU/pediatric intensive care units who received a total of 113,773 TPN orders recorded between January 2011 and January 2022. These TPN data represent best practice in the field and come as close as possible to representing the gold standard given the knowledge currently available to the field. All TPN orders were developed by a multidisciplinary healthcare team. Each order is placed by a neonatologist, house staff resident or neonatal nurse practitioner, and is then checked by a NICU clinical pharmacist and a NICU dietitian. They also use modular tools, such as those embedded in Epic, for automated calculation to ensure accurate molecular dosages and appropriate physicochemical properties. Of the 6,991 patients identified, 5,913 were retained as they received their first TPN within the first 2 years of delivery, which resulted in 79,790 TPN orders. Sex and gender were not considered in the study design as this is a retrospective study of all available patients that fit the inclusion criteria. Of these, 68% of the prescriptions were TPN with no enteral feeds, and the rest were partial PN with some amounts of enteral feed. These data include patients receiving both total and partial parenteral nutrition as they transition to increased enteral feeding while preparing for discharge from the NICU.

In addition, an independent validation cohort was obtained with EHRs from the UCSF. All 3,417 patients who received their first TPN within the first 2 years of birth between October 2012 and January 2024 in the UCSF EHR Database were included (63,273 TPN orders). The curated UCSF EHR data contain demographics, specific laboratory measurements and TPN data. This study was approved by the Institutional Review Board of UCSF (reference no. 17-22929) with consent waived.

Data processing and packages

At Stanford, gestational age at delivery and birth weight were extracted from clinical notes in the newborns' EHRs. Free text in clinical notes was searched systematically using regular expressions. When several clinical notes were available for the same newborn and values were discordant, the most commonly occurring value was retained or the average across all the different values if two or more values appeared with the same frequency. To check the accuracy of the information extracted from clinical notes compared to manual chart review, in our previous study, we randomly sampled 40 newborns. Their gestational age at delivery and birth weight were extracted by reviewing electronic charts manually. Comparison of these systematically extracted data with those extracted manually from clinical notes showed perfect accuracy[81].

For each newborn, their entire medical history available in the EHRs corresponding to their days on TPN was extracted. This consisted of all conditions, observations, medications and procedures, recorded under the patient identification number. Different types of records were as follows: (1) conditions (presence of a disease or medical condition); (2) observations (observed clinical sequelae obtained as part of the medical history); (3) medications (utilization of any prescribed and over-the-counter medicines, vaccines and large-molecule biologic therapies) and (4) procedures (records of activities or processes ordered by or carried out by a healthcare provider on the patient for a diagnostic or therapeutic purpose). Conditions, observations, medications and procedures were organized by patient and date and time of the day at which they were entered into the EHR system.

About 5% of the gestational age and birth weight data were missing, and they were imputed using missing forest based on the phenotypic data. Other missing laboratory values were imputed using the patient's latest past known values, if any, otherwise mean values of the cohort were used. For the EHR data, each procedure, condition, drug and observation were represented in a binary format. All analyses were performed using Python v.3.10, and the deep-learning models were implemented using either pyTorch (v.2.0.1) or Tensorflow (v.2.11.0) on a linux-based server with Intel(R) Xeon(R) Gold 6330 and NVIDIA A40 equipped. Data processing and analyses were performed using Scipy (v.1.11.1), Numpy (v.1.24.4), scikit-learn (v.1.2.2), Pandas (v.2.2.2), statsmodel (v.0.14.0), sksurv (v.0.21), lifelines (v.0.28.0), missForst (v.3.1.3) and forestplot (v.0.4.1).

TPN2.0 by variational neural network and iterative clustering

Our VNN was an adaptation of a variational autoencoder (AE)[83]—a type of generative deep-learning model, to better suit the task of predicting TPN compositions from the inputs while maintaining the ability to generate latent representation of the patient via the bottleneck layer. The VNN bottleneck layer contained 16 nodes, which generated the latent space for subsequent clustering.

The inputs of the model were composed of two parts. The first part is a 32-dimensional latent representation obtained from a multimodal AE, where the first two layers were modality specific, then connected at the final layer before the bottleneck. The decoder mirrored this structure. The input and output include all EHR data, which consisted of four modalities: 1,625 observations; 1,037 medications; 5,607 conditions and 6,230 procedure features. The second part of the VNN inputs was the primary data physicians typically look at to determine TPN dosages. These include basic characteristics (sex, race, days since birth, gestational age, birth weight, dosing weight and cholestasis status), whole blood serum values (alanine transaminase, albumin, alkaline phosphatase, aspartate transferase, blood urea nitrogen, calcium, chloride, creatinine, glucose, ionized calcium, magnesium, phosphate, potassium, sodium, triglyceride) and feed information (total fluid volume, enteral fluid volume, TPN volume, infusion rate, protocol (neonatal versus pediatric), feed line (central versus peripheral), days on TPN, hours of TPN and fat product type). The second part of the input data was not compressed and was fed into the VNN explicitly (Fig. 2a). The outputs of the model are the TPN dosages which include fat (g kg^{-1}), amino acids (g kg^{-1}), dextrose (%), acetate (mEq kg^{-1}), calcium gluconate (mg kg^{-1}), copper (mcg kg^{-1}), famotidine (mg kg^{-1}), heparin (units ml^{-1}), levocarnitine (mg kg^{-1}), magnesium (mEq kg^{-1}), multivitamins (ml kg^{-1}), phosphate (mmol kg^{-1}), potassium (mEq kg^{-1}), selenium (mcg kg^{-1}), sodium (mEq kg^{-1}) and zinc (mcg kg^{-1}).

The Stanford cohort data were separated randomly during the twofold cross-validation, where each fold is structured as 40% training, 10% validation and 50% test set. The separation was done at the patient level, that is, no different TPN orders from the same patient were in both the training and test sets. The model was trained using a batch size of 128 and optimized using an adaptive moment estimation (Adam) optimizer to minimize the average root mean squared error (r.m.s.e.) across all TPN components in the training set. The loss in the validation set was used to determine the stopping point of the training process. The model was then evaluated on the unseen test set using Pearson's correlation between each pair of actual an actual prescription and TPN2.0 as the metric. We also reported performance compared to human experts (Fig. 3b). To first evaluate human performance (the gold standard), we calculated the Euclidean distance between a given prescription and another actual prescription from a patient with similar characteristics. A patient with similar characteristics was identified by randomly selecting one whose laboratory serum values and baseline characteristics were all within a 10% range of difference from the original patient's values. To evaluate TPN2.0 performance, a similar approach was taken, but TPN2.0 recommendations of the matched patient were used to calculate Euclidean distance instead. The distance

performance of human and TPN2.0 was then compared using normalized distance and correlation metrics.

The model's performance was also evaluated using an independent cohort from UCSF. We trained a separate VNN model on a subset of features that overlaps both Stanford and UCSF data. This VNN version's input did not include aspartate transferase, blood urea nitrogen, creatinine, enteral information, fat product, hours of TPN, infusion rate, ionized calcium or total fluid information. It was trained and evaluated at Stanford using the same protocol as described above, and then evaluated on the UCSF cohort without any retraining.

After training and validating the VNN model, we employed iterative semisupervised clustering on the training set to condense the number of predicted components into a finite set of formulas. For each combination of protocol and feeding line, for example, neonatal central line, the 16-dimension VNN latent representations were grouped into 100 clusters using *K*-means clustering. The Euclidean pairwise distances between these cluster centroids were calculated and sorted. We then selected the 30 pairs with the smallest distances. For each pair, we assessed the changes in correlations if the pair were to be merged. The correlations were between the compositions from the assigned clusters and the actual prescriptions. The pair that resulted in the smallest change in correlation was merged. Essentially, we first overclustered into 100 clusters and iteratively merged the pair of clusters that minimized the reduction in correlation to the actual prescriptions. This process repeated until we achieved the desired number of clusters. The final model was then applied to the test set without retraining.

Visualization of high-dimensional latent space

The 16-dimensions latent space from the VNN was reduced to 2D for visualization using uniform manifold approximation and projection. We randomly sampled 20,000 data points for the visualization. Uniform manifold approximation and projection was initialized with the following parameters: n_neighbors = 20, min_dist = 0.001, metric = 'chebyshev' and spread = 3. The resulting 2D datapoints were colored according to the TPN2.0 predicted clusters.

Evaluation of consistency in actual prescriptions and TPN2.0

To determine whether the variability of TPN2.0 was lower than that of actual prescriptions (Extended Data Fig. 3), we first needed to group similar patients together to be able to determine accuracy and consistency within each group. An AE with a 16-dimensional latent space was constructed to encode and decode the input data. The encoder and decoder used neural network layers with rectified linear units and sigmoid activations, respectively. Afterward, the AE was trained using mean squared error loss and the Adam optimizer. Post-training, the latent space representation was extracted for clustering. *K*-means clustering was performed with a range of 5 to 50 clusters. The optimal number of clusters was determined by two methods: (1) silhouette scores and (2) plotting the within-cluster sum of squares and using the KneeLocator method (Supplementary Fig. 11). We used the highest optimal number of clusters between these methods—29 clusters—to take a more conservative approach and improve the assessment of physician prescription variance.

To ensure that similar patients were clustered together and that the reduction in variance was due to accurate TPN2.0 clustering and not merely the result of reducing the number of possible formulas, a randomized clustering test was conducted. TPN2.0 clusters were shuffled 1,000 times and the weighted mean variances were recalculated. In all iterations, the actual TPN2.0 clusters had lower variance ($\sigma^2 = 0.76$) than the random cluster ($\sigma^2 = 0.99$), indicating the decrease in variance was due to accurate clustering.

Blinded study of TPN2.0

We conducted a blinded study to evaluate the healthcare team's preference between actual prescriptions and TPN2.0. In this study, we recruited members of the healthcare team who were typically involved in TPN prescription to rate different TPN solutions from 0 to 100 when presented with a patient profile. Up to 30 patients were randomly selected from Stanford historical records. For each patient, the three TPN solutions were presented: the actual prescription given to the patient, the TPN2.0 cluster composition and a random TPN. The random TPN composition was drawn randomly from other actual prescriptions in the dataset. The team was blinded to the labels of the three TPN solutions; they also did not have access to the previous days' TPN records. Other than these, the team was allowed to access any EHR of the patient. The healthcare team participating in the study included pediatric residents, neonatology fellows, neonatologists and pharmacists. This generated a total of 192 comparisons.

A similar blinded study was conducted with a physician to identify whether the distance between TPN2.0 and the actual prescription was confounded by the physician adjusting prescription composition in expectation of a certain diagnosis. In this study, only ten TPN from different patients were used. These patients were the ones who exhibited the highest distances between TPN2.0 and actual prescriptions before cholestasis diagnosis.

Outcome extraction and analysis against TPN2.0

We collected 17 neonatal outcomes of interest, including anemia, BPD, candidiasis, cerebral palsy, cholestasis, CHD, mortality, intraventricular hemorrhage, jaundice, NEC, patent ductus arteriosus, periventricular leukomalacia, pulmonary hemorrhage, pulmonary hypertension, respiratory distress syndrome, retinopathy of prematurity and sepsis. The selected clinical morbidities are among the most commonly observed in patients in the NICU setting[84–90], especially in preterm infants. Although some, such as NEC, are more commonly seen in preterm infants, others like BPD and sepsis affect both term and preterm infants, and may be influenced by various factors, including the use of TPN[91]. Many of these conditions are interconnected. For example, conditions like NEC, which is linked to intestinal injury, can occur in term infants, particularly those with underlying conditions such as hypoxia in congenital cardiac diseases[92]. Although these morbidities are well recognized in NICU care, many of them remain underinvestigated, particularly in the context of emerging tools such as ML. This analysis, therefore, includes these conditions to provide a comprehensive evaluation of the potential risks that influence neonatal health outcomes, not solely limited to known risks associated with TPN use.

Outcomes were considered to be present or absent based on neonates' health record data. The outcome data were extracted from the newborn's medical history at any timepoint since their birth up until 3 months after their last day on TPN. Grouped disorders (for example, other central nervous system disorders or liver diseases) affecting the same organ system were excluded due to nonspecificity and confounding potentials. The list of SNOMED concept codes used to identify the presence/absence of each outcome is reported in Supplementary Table 3. A blinded manual chart review of the outcomes considered for 30 randomly sampled preterm newborns was conducted in our previous study to assess the sensitivity and specificity of the definitions used[93]. Overall sensitivity and specificity (across all outcomes) were 97.5% and 98.8%, respectively, showing high concordance between manual chart review adjudication and EHR-based definitions.

To evaluate whether the rates of adverse outcomes changed according to the similarity between TPN orders and TPN2.0 recommendations, we calculated the distance between the composition of each order and the TPN2.0 recommendations. The distance represents how similar the two were, where shorter distances indicated higher similarity. If the rates of outcomes changed with distance, that means prescribing more (or less) similar to TPN2.0 was associated with outcomes. The distance was calculated using the Manhattan distance

across all standardized nutrient values. Three analyses were used to evaluate the distance versus outcome rates relationship: area under the receiver operating characteristic (AUROC), OR and survival analyses. The AUROC metric indicated if an increase in distance was associated with higher rates of adverse outcomes. In the OR analysis, we split the data into case and control groups by patients with less than the 80th percentile distance from TPN2.0 being the control and those with more being the case. We then calculated the OR for each outcome between these two groups. Outcomes with a significant OR ($P < 0.05$; Woolf test) and an AUROC > 0.6 suggested a trend where higher distances predicted higher risks of adverse outcomes (Supplementary Fig. 12) and were presented in the main findings. For the survival analysis, we replaced the commonly used time component with the Manhattan distance. To determine whether there was a difference in the rate of adverse outcomes between those who were very close to TPN2.0 and those who were very far, we split the case and control groups into those more than the 80th percentile away from TPN2.0 and those within the 20th percentile distance from TPN2.0, respectively. The P value to determine the rate differences between the two groups was obtained from a multivariate log rank test[94]. Note that, in all analyses, we used only the information from up to 3 months before the day of diagnosis and removed any cases with CHD to limit confounding potentials. For the analysis of NEC, we limited it to only prescriptions within the first 6 months of life to limit any false discovery, as the disease happens only early in life.

Development of deep sequential models for TPN

Sequential models could leverage the longitudinal nature of TPN data to increase model performance. Three deep-learning variants were investigated: long short-term memory (LSTM)[95], temporal Kolmogorov–Arnold networks (KAN)[96,97] and transformer[98]. LSTM is a variant of recurrent neural network architecture that allows the model to learn from sequential data. A sequence-to-sequence LSTM model was constructed such that, for each input, there was an output corresponding to that day. The hidden state from the timepoint t is then recycled as part of the inputs for timepoint $t+1$ for predicting the next outputs. Our LSTM model consisted of a 256-node LSTM layer followed by a fully connected layer with sequence padding to address variable lengths. Afterward, a sequence-to-sequence KAN model was also investigated. KAN architecture is a recently developed alternative to multilayer perceptron, which is what LSTM, transformer and most other neural network models are based on, that showed superior performance in many applications and features, such as stability and interpretability. Essentially, KANs have learnable activation functions on edges, while multilayer perceptrons have fixed activation functions on nodes. KANs also have no linear weights, with every weight parameter replaced by a univariate function such as a spline. Similar to LSTM, our temporal KAN model consisted of a 256-node recurrent neural network-like KAN layer followed by a fully connected layer.

Transformer architecture is a neural network that learns the context of sequential data (encoding) and generates new data from it (decoding) using an attention mechanism to focus on relevant parts of the sequence. To predict the TPN dosage for timepoint t, the encoder was fed input from timepoint t and the decoder was fed the TPN dosage from timepoint $t-1$ (Fig. 6a). We built a transformer model with a layer of 256-node encoder and decoder with eight attention heads. The model was first pretrained using the teacher forcing method[98]—a common training technique where the transformer's decoder was fed the correct output (actual prescriptions) during training. Similar to VNN, the loss function to minimize was r.m.s.e. and the optimizer was Adam. For the PI-transformer, we imposed additional losses with regard to conditions and boundaries described in Supplementary Table 2. Note that the solubility was calculated using calcium phosphate curve approximation equations[99]. The model was then trained without the teacher forcing or using the actual prescriptions as input, that is, autoregressively generated the prediction of timepoint t and fed it as the decoder's input for prediction of timepoint $t+1$. Once trained, a clustering layer[100] was appended to the previous prediction output layer. The clustering layer was initiated using K-means cluster centroid. The model was then trained to minimize an equal mix of clustering loss[100] and the r.m.s.e. loss from predicted outputs before clustering until the cluster assignment converged. We found this method performed better in the TPN dataset context.

Inference was used to determine the transformer model's performance measured by Pearson's correlation. By default, inference makes predictions autoregressively, which is similar to the way we trained the model but without updating the model's weights. The flexibility of the inference method also allowed us to investigate its next-day performance in a scenario where TPN2.0 recommendations were to be modified today. To conduct a study mimicking this scenario, the Euclidean distances between the model predictions and the actual prescriptions were calculated in the training set. A distance cutoff was chosen based on the desired percentile, that is, a cutoff at the 80th percentile distance if we are studying 20% intervention. During inference, any prediction with a distance over the cutoff was replaced by the actual prescription as the decoder's input. This mimics a scenario where the predicted TPN composition was too different from the providers' expectations; hence, they modified it. Note that when the actual prescription was used as the decoder's input, we did not take that prediction instance for the calculation of the model performance.

Feature attribution was performed using Gradient SHAP—a variant of SHapley Additive exPlanations designed for deep-learning models[101]. The method leverages the gradients of the model's outputs with respect to its inputs to estimate feature importance, enabling the decomposition of predictions into contributions from individual features. To simplify the analysis, the longitudinal data was flattened into a cross-sectional format. We summarized the training dataset into 100 representative background samples using K-means clustering. For each cluster assignment probability, SHAP values were computed for 7,500 randomly selected test samples.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Stanford University and University of California, San Francisco are the custodians of their own individual patient-level protected information. Access to these data is subject to approval from the respective institutions. To request data access, researchers or institutions should contact N.A. at https://nalab.stanford.edu/. The request will further require approval from Stanford University Privacy Office and the Institutional Review Boards (IRBs). Requests will be reviewed on a rolling basis and timeline is dependent on the IRB. Data will be for research purposes only and must comply with ethical guidelines and institutional policies on data privacy.

Code availability

A repository with the original Python code used to train the VNN and the transformer model is available on GitHub at https://github.com/tpjoe/TPN2.0.

References

83. Kingma, D. P. Auto-encoding variational Bayes. Preprint at https://arxiv.org/abs/1312.6114v11 (2013).
84. Barnette, A. R., Myers, B. J., Berg, C. S. & Inder, T. E. Sodium intake and intraventricular hemorrhage in the preterm infant. *Ann. Neurol.* **67**, 817–823 (2010).
85. Toya, Y. et al. Effects of total parenteral nutrition on serum osmolality and patent ductus arteriosus. *Cureus* **16**, e64196 (2024).

86. Akın, M. Ş. & Yiğit, Ş. The effects of early nutritional contents in premature infants on the development and severity of retinopathy: a retrospective case-control study. *Trends Pediatr.* **4**, 238–246 (2023).
87. Cannon, R. A. et al. Home parenteral nutrition in infants. *J. Pediatr.* **96**, 1098–1104 (1980).
88. Yeo, S. L. NICU update: State of the science of NEC. *J. Perinat. Neonatal Nurs.* **20**, 46 (2006).
89. Synnes, A. R. et al. Neonatal intensive care unit characteristics affect the incidence of severe intraventricular hemorrhage. *Medical Care* **44**, 754 (2006).
90. Leng, Y. et al. The treatment and risk factors of retinopathy of prematurity in neonatal intensive care units. *BMC Ophthalmol.* **18**, 301 (2018).
91. Rodriguez, S., de la Cruz, D. & Neu, J. Nutrition strategies to prevent short-term adverse outcomes in preterm neonates. *BMJ Nutr. Prev. Health* https://doi.org/10.1136/bmjnph-2023-00080 (2024).
92. Noerr, B. Part 1. Current controversies in the understanding of necrotizing enterocolitis. *Adv. Neonatal Care* **3**, 107–120 (2003).
93. De Francesco, D. et al. Data-driven longitudinal characterization of neonatal health and morbidity. *Sci. Transl. Med.* **15**, eadc9854 (2023).
94. Davidson-Pilon, C. lifelines: survival analysis in Python. *J. Open Source Softw* **4**, 1317 (2019).
95. Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural Comput.* **9**, 1735–1780 (1997).
96. Liu, Z., et al. Kan: Kolmogorov-Arnold networks. Preprint at https://arxiv.org/abs/2404.19756v5 (2024).
97. Genet, R. & Inzirillo, H. TKAN: Temporal Kolmogorov-Arnold networks. Preprint at https://arxiv.org/abs/2405.07344v3 (2024).
98. Vaswani, A. Attention is all you need. In *Proc. Advances in Neural Information Processing Systems 30 (NIPS 2017)* (eds Guyon, I. et al.) 5999–6009 (Neural Information Processing Systems, 2017).
99. Anderson, C., Eggert, L., Fitzgerald, K., Jackson, D. & Farr, F. Calcium and phosphate solubility curve equation for determining precipitation limits in compounding parenteral nutrition. *Hosp. Pharm.* **57**, 779–785 (2022).
100. Guo, X., Gao, L., Liu, X. & Yin, J. Improved deep embedded clustering with local structure preservation. In *Proc. Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17)* (ed Sierra, C.) 1753–1759 (International Joint Conferences on Artificial Intelligence Organization, 2017).
101. Lundberg, S. A unified approach to interpreting model predictions. Preprint at https://arxiv.org/abs/1705.07874v2 (2017).

Acknowledgements

We thank N. Shah, N. Pageler, N. Mackenzie and the NeoMIND-AI group for contributions to the project or the infrastructure that enabled this work. This research used data or services provided by STARR, 'STAnford medicine Research data Repository', a clinical data warehouse containing live Epic data from Stanford Health Care, the Lucile Packard Children's Hospital and other auxiliary data from hospital applications such as Infusion Studio. The STARR platform is developed and operated by the Stanford Medicine Research IT team and is made possible by Stanford School of Medicine Research Office. We acknowledge the use of the UCSF Information Commons computational research platform, developed and supported by UCSF Bakar Computational Health Sciences Institute in collaboration with IT Academic Research Services, Center for Intelligent Imaging Computational Core and CTSI Research Technology Program. This work was supported by NIH grant nos. R42HD115517, R35GM138353 and UL1TR001872, the Burroughs Wellcome Fund (1019816), the March of Dimes, the Alfred E. Mann Foundation, the Stanford Maternal and Child Health Research Institute through Stanford's SPARK Translational Research Program, Stanford High Impact Technology (HIT) Fund and Stanford Biodesign. This project was also supported by NCATS through the UCSF Clinical and Translational Science Institute (CTSI). This work is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies.

Author contributions

T.P., M.G., J.D.R. and T. T. O. contributed equally to the primary data aggregation, design and all experiments conducted in this study. T.P. and D.D.F. led the algorithm development and analysis, with critical contributions from S.J.M., S.M.R., C.E., P.C., F.X., E.B., Y.K., D.S., L.W.Y.S., C.-H.S., S.S., L.K.Y. and Y.T. Data curation and technical support for independent validation was assisted by J.M.C., B.O. and J.R. The healthcare team contributing to the blinded study, reviewing cases and algorithm designs include J.D.R., T.N., J.A.S., S.R., M.Y.S., F.M., H.G., J.F., C.P.V. and L.A.R. S.L., R.J.W., B.G., M.S.A., T.J.M., J.A.K., R.L.K., G.M.S., K.G.S., J.F., V.C., S.G. and D.K.S. provided domain-specific expertise and insights for study methodology as well as reviewing the manuscript to ensure clinical relevance and accuracy. M.S. led the supervision of the independent model validation. L.S.P. provided strategic guidance and facilitate blinded studies. N.A. supervised the research, secured funding and provided strategic guidance.

Competing interests

The methods described in this manuscript are covered in the US provisional Patent 63/268,689 (WO2022256850A1; 'Systems and methods to assess neonatal health risk and uses thereof') approved in 2022. T.P. is a cofounder of Takeoff41. S.M. is a paid consultant for Danaher and Longitude Capital and receives a paid fellowship from Nucleate. J.H.F. is an advisor to Vitara, OvaryIt, Keriton, EmpoHealth, and Avanos; the consulting medical director of Novonate; and a cofounder for EMME. K.G.S. is a consultant for Avexegen Therapeutics, Infinant Health, mProbe and Mission Biocapital. M.S.A. is a member of the Scientific Advisory Board of Cytonics Inc. and AfaSci Research Laboratories and is a paid consultant for Syneos Health. D.K.S. is a member of the Clinical Advisory Board of Maternica Therapeutics. M.S. is a member of the Scientific Advisory Board of Exagen and Aria Pharmaceuticals and is a shareholder at Somnics. N.A. is a member of the Scientific Advisory Boards of January AI, Parallel Bio and WellSim Biomedical Technologies, is a cofounder of Takeoff41 and is a paid consultant for MaraBio Systems. The other authors declare no competing interests.

Additional information

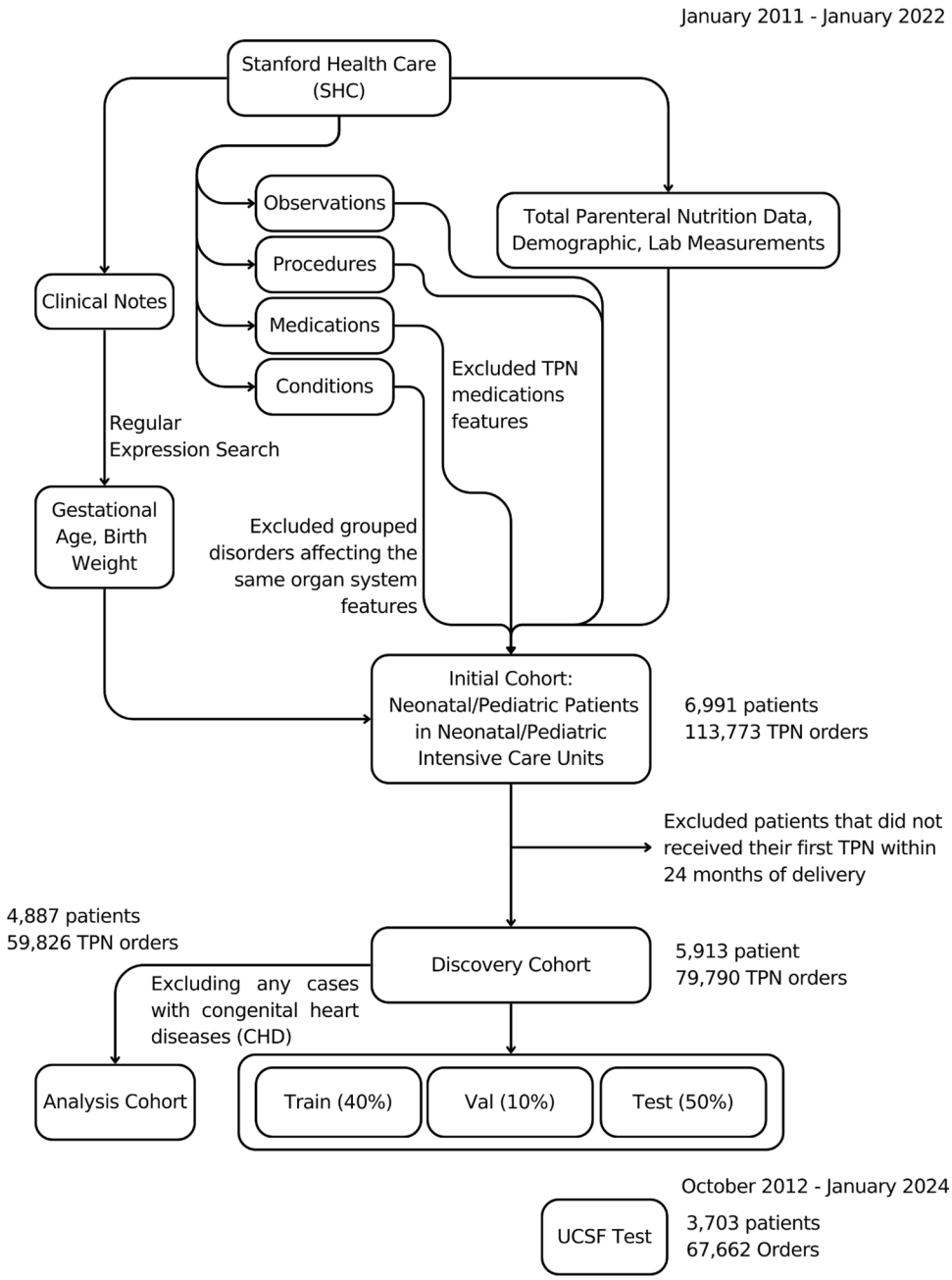
Extended data is available for this paper at https://doi.org/10.1038/s41591-025-03601-1.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41591-025-03601-1.

Correspondence and requests for materials should be addressed to Nima Aghaeepour.

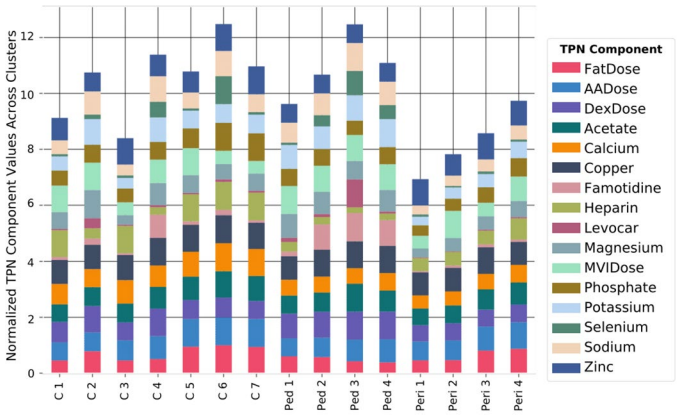
Peer review information *Nature Medicine* thanks Kensaku Kawamoto, Neena Modi and Ola Saugstad for their contribution to the peer review of this work. Primary Handling Editor: Lorenzo Righetto, in collaboration with the *Nature Medicine* team.

https://doi.org/10.1038/s41591-025-03601-1



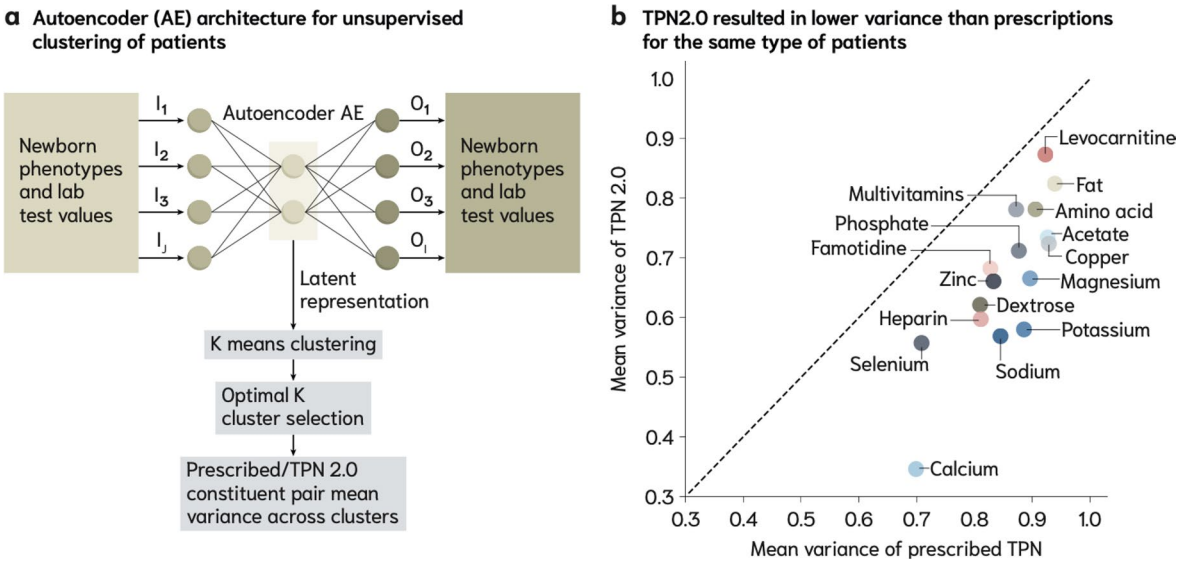
Extended Data Fig. 1 | Flowchart diagram of the dataset generation process. The data were aggregated from the EHRs at Stanford Health Care. The linkage of the two datasets allowed for a combination of nutritional data, phenotypic traits, and long-term outcome data, among others. All EHR data were mapped to the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) version 5.3.1, which included patient observations, procedures, medications, and conditions. Gestational age at delivery and birth weight were extracted from clinical notes in the newborns' EHRs using regular expressions. For each newborn, their entire medical history available in the EHR corresponding to their days on TPN was extracted, including all conditions, observations, medications, and procedures, while excluding TPN medication records to avoid potential data leakage. Conditions, observations, medications, and procedures were organized by patient, date, and time of entry into the EHR system. Conditions that affect multiple organ systems were excluded. Initially, the Stanford cohort included 6,991 neonatal/pediatric patients in neonatal/pediatric intensive care units with 113,773 TPN orders recorded between January 2011 and January 2022. Of these, 5,913 patients were retained as they received their first TPN within the first 2 years of delivery, resulting in 79,790 TPN orders. EHR data were represented in a binary format for each condition, drug, procedure, and observation. The patients with congenital heart diseases were dropped for the TPN2.0 outcome comparison analysis. In addition, an independent external validation cohort was obtained from EHRs from the UCSF Hospital and Clinics and the Benioff Children's Hospital. The UCSF EHR database contains demographics, specific lab measurements, and TPN data. This included all 3,417 patients who received their first TPN within the first 24 months of birth between October 2012 and January 2024 in the UCSF EHR Database, totaling 63,273 TPN orders.

https://doi.org/10.1038/s41591-025-03601-1



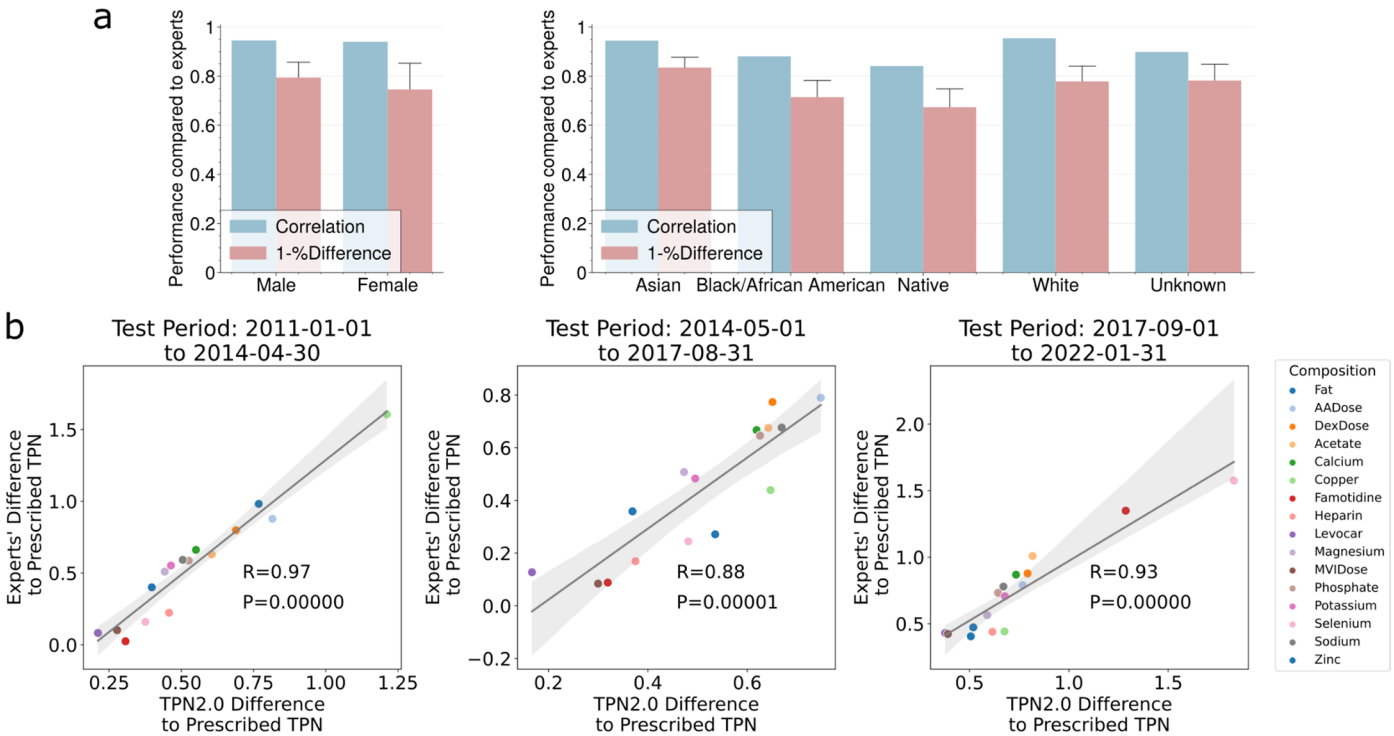
Extended Data Fig. 2 | The composition of TPN2.0 cluster representatives. Composition of the 15 clusters identified by iterative clustering of TPN2.0's latent representations. The ratio values (y-axis) are obtained by normalizing each component across clusters for visualization purposes only. The cluster types are divided into 7 central and 4 peripheral lines from the neonatal protocol, and 4 for the pediatric protocol.

https://doi.org/10.1038/s41591-025-03601-1



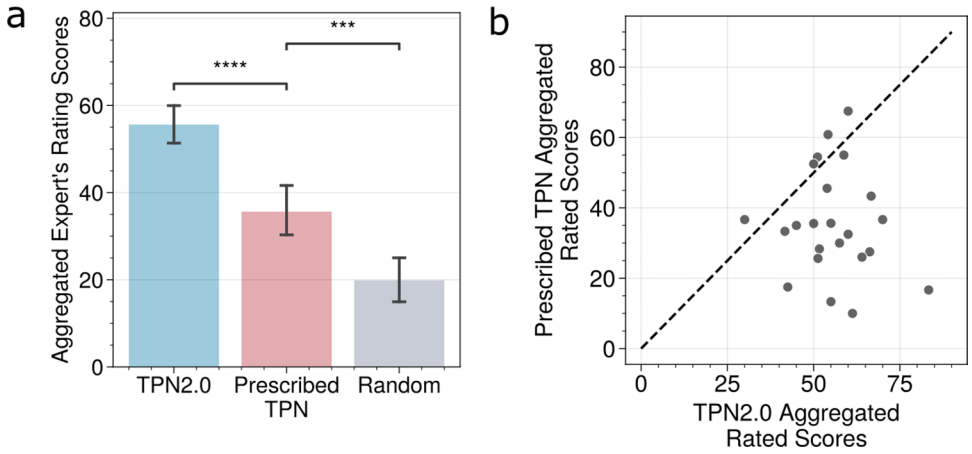
Extended Data Fig. 3 | TPN2.0 exhibited improved unexplained variability than current best practices. a, An AE model was used to generate a latent representation of the patients, followed by clustering, in order to identify homogeneous patient groups. The model was applied to patient characteristics and lab test values. During this process, a compressed latent representation of the input is extracted. This representation is then grouped by K-means clustering. The optimal number of clusters are determined by using Silhouette scores and Within-Cluster Sum of Squares (WCSS) using the KneeLocator method. Subsequently, the variance of each TPN component within each cluster was calculated for both TPN2.0 and the actual prescriptions. **b**, The mean variance across all clusters (weighted by cluster size) is visualized. The scatter plot shows lower mean variances within the patients in the same group for all components in TPN2.0 compared to current best practice.

https://doi.org/10.1038/s41591-025-03601-1



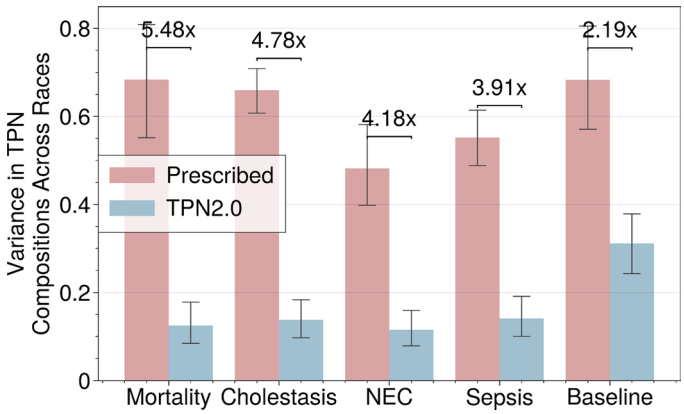
Extended Data Fig. 4 | TPN2.0 performance generalizes over different sexes, races, and periods. a, Stratification of the model performance as presented in Fig. 3b by sexes and races indicates that the high performance applies across different subpopulations. The corresponding powers of the stratified analyses are 1 for both male and female at the significance level (α) of 0.0001, and 1 for Asian (α = 0.0001), 0.93 for African American (α = 0.001), 0.94 for Native (α = 0.01), 1 for white (α = 0.0001), and 0.89 for race unknown (α = 0.0001). The race 'Native' includes both Native Hawaiian or Other Pacific Islander and American Indian or Alaska Native due to the insufficient numbers of population from these races in our cohort. Refer to Supplementary Table 1 for the number of data points in each group. Data are presented as mean values +/− SEM. **b**, The model performance with experts as described in Fig. 3c with these results obtained from a time-based cross validation instead of a random train-test split. In each cross-validation, the model was trained on data from a time period that did not overlap with the test period. For example, in the last plot, the model is trained on data from January 1st 2011 to August 31st 2017, and it is tested on TPN orders from September 1st 2017 to January 31st 2022. The high correlation across all periods suggests that potential changes in clinical guidelines from different periods did not meaningfully impact the model performance.

https://doi.org/10.1038/s41591-025-03601-1



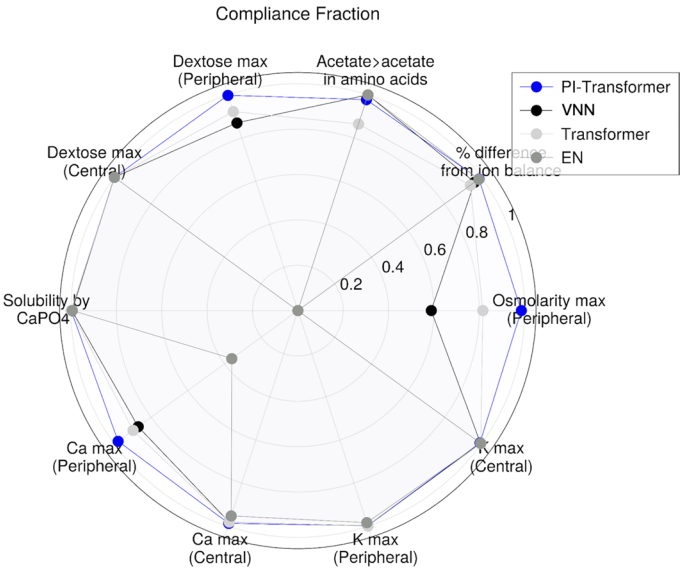
Extended Data Fig. 5 | TPN2.0 outperforms current best-practices, even when compared to a mixture-of-experts approach derived from individual expert ratings aggregated at the patient level. a, Instead of analyzing rating scores from an individual expert for each patient, we evaluate the blinded study by aggregating scores at the patient level, and only include those rated by at least 3 experts. This resulted in a total of 23 comparison pairs. The bar chart showing the aggregated rating scores for TPN2.0, prescribed TPN, and random TPN formulations still shows that TPN2.0 receives the highest ratings, significantly outperforming both the prescribed TPN from best practice and random formulations. Error bars represent standard errors across the aggregated scores. Data are presented as mean values +/− SEM. **b**, Furthermore, looking at the individual patient's averaged rating score, each represented by a dot in the scatter plot, TPN2.0 was rated either about the same as prescribed TPN, or in most cases, higher than them. These results support the preference for TPN2.0, even when evaluated using a method that more closely reflects clinical best practices.

https://doi.org/10.1038/s41591-025-03601-1



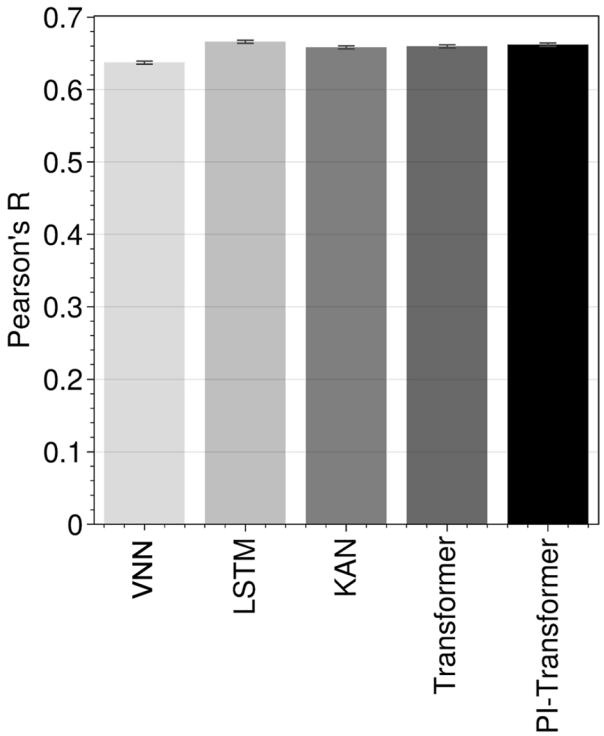
Extended Data Fig. 6 | TPN2.0 reduces race-specific variance, particularly in populations with adverse outcomes. The variance of the TPN composition values across races is calculated for both TPN2.0 and actual prescriptions. The calculation is stratified for patients with specific adverse outcomes and for those who were not diagnosed with any of the 16 adverse outcomes or congenital heart diseases (baseline) listed in Supplementary Table 3. The results indicate that actual prescriptions are not only associated with higher variance across races in baseline patients compared to TPN2.0 (~2x), but more so (>~4x) in patients with adverse outcomes. Refer to Supplementary Table 1 for the number of data points in each group. Data are presented as mean values +/− SEM.

https://doi.org/10.1038/s41591-025-03601-1



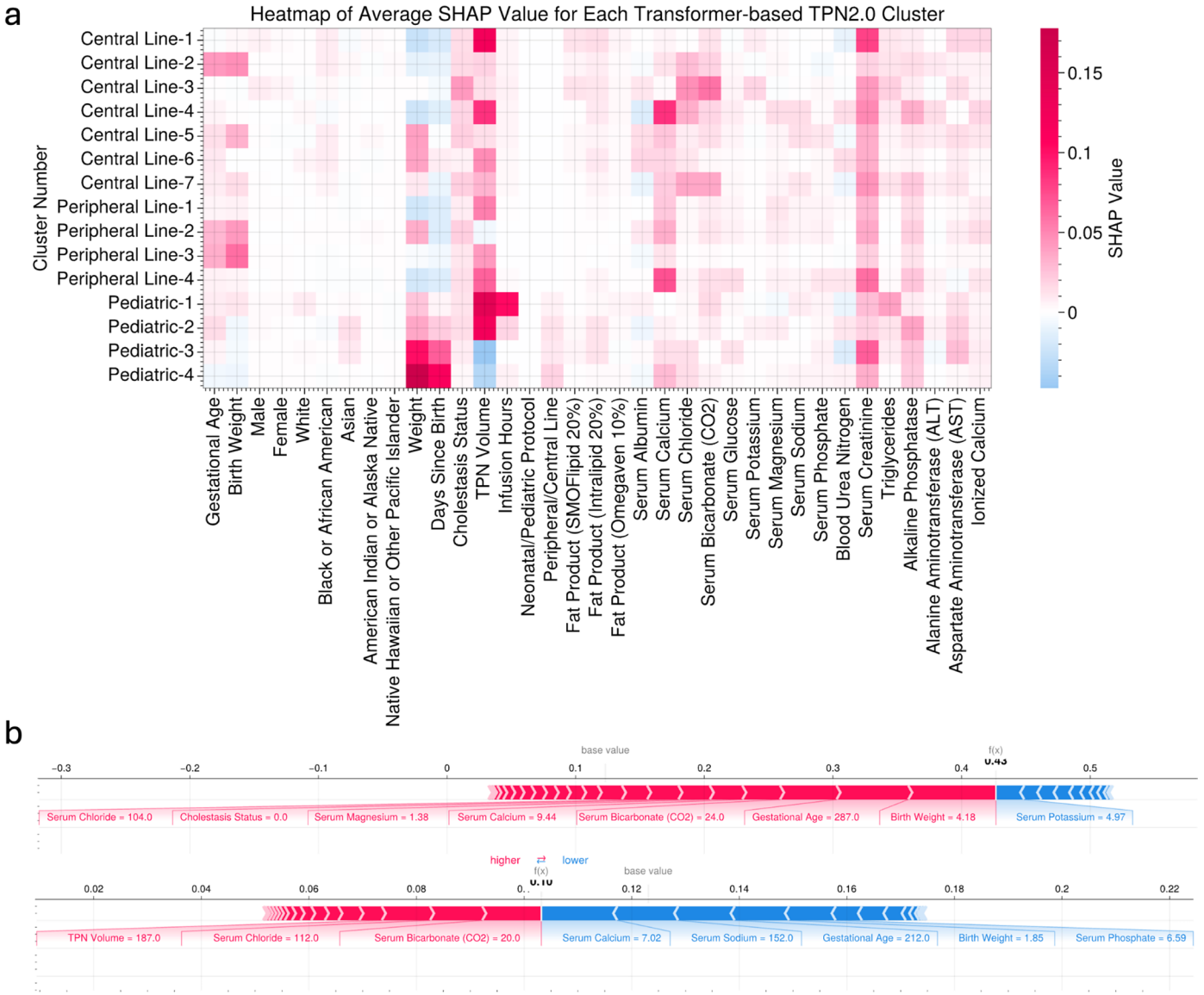
Extended Data Fig. 7 | Fraction of TPN compositions that adhere to each criterion of pharmacist/clinical guidelines or physical expectations. The limits for these criteria are listed in Supplementary Table 2. Here, physics-informed (PI) transformer vastly outperforms normal transformer, VNN, and baseline Elastic Net.

https://doi.org/10.1038/s41591-025-03601-1



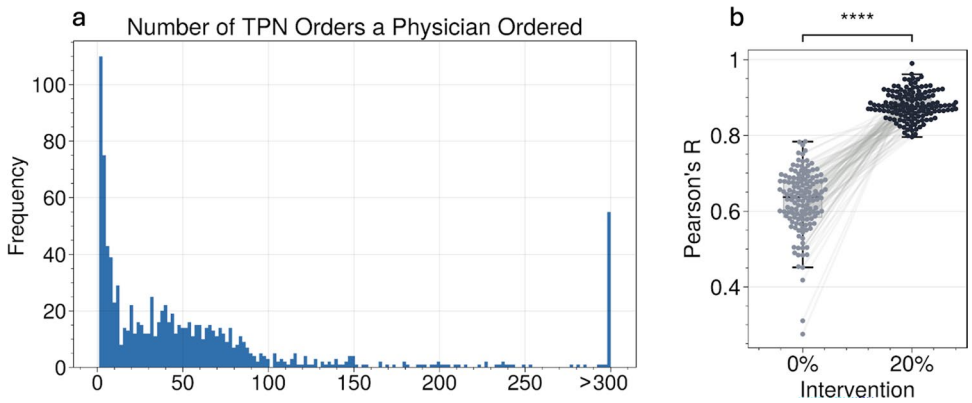
Extended Data Fig. 8 | Physics-informed (PI) transformer accommodates safety criterion while maintaining accuracy comparable to other algorithms. Comparison of the performance (n = 79,790 TPN from 5,913 patients) of different deep learning architectures, including VNN, Long Short-Term Memory (LSTM), Kolmogorov–Arnold Networks (KAN), transformers, and PI-transformer, all showing similar performance. The performance is a Pearson's R of TPN2.0 composition vs. the actual prescription composition. Data are presented as mean values +/− SEM.

https://doi.org/10.1038/s41591-025-03601-1



Extended Data Fig. 9 | Feature attribution analysis reveals the features driving the predictions of the transformer-based TPN2.0. a, A heatmap of average gradient-based SHAP (SHapley Additive exPlanations) values visualizes global feature importance for each selected feature and cluster. The SHAP values are derived from 7,500 randomly and uniformly selected TPN orders. The clusters are organized into neonatal central line, peripheral, and pediatric protocol clusters. For each cluster, only SHAP values from samples assigned to that cluster are considered. Color intensity represents the magnitude of SHAP values, with red indicating positive contributions and blue indicating negative contributions to cluster assignment. For example, assignment to cluster 1 of the central line is heavily influenced by serum creatinine levels, while cluster 4 relies more on serum calcium levels. Importantly, the model does not heavily depend on race or sex for cluster assignments, minimizing demographic-based biases. **b**, Force plots demonstrate how features drive cluster assignments at a local, patient level. The top plot shows a sample assigned to cluster 2 of the central line, where features like birth weight, gestational age, and serum bicarbonate exert strong positive influences, increasing the assignment probability to 0.43 (~3.5x the base value of 0.123). In contrast, the bottom plot shows a sample with a reduced probability of 0.10 for cluster 2 due to negative contributions from low serum calcium, serum sodium, gestational age, and birth weight. These examples highlight the complexity of the model's predictions at an individual level, revealing feature-specific contributions that may not be fully captured by global explanations.

https://doi.org/10.1038/s41591-025-03601-1



Extended Data Fig. 10 | TPN2.0 with physician-in-the-loop recommendations leads to performance improvement for every physician. a, The number of TPN orders prescribed by each physician in the dataset. **b**, Collaboration with TPN2.0 leads to improved performance for every individual physician. Figure 6c previously shows that the model's performance improves with increasing levels of physician intervention, where physicians adjust TPN2.0's recommendations in a simulated scenario. Here, we show that the improvement applies to every single individual. Each dot in the pair plot represents the average correlation of TPN2.0 to all actual prescription by a physician.

nature portfolio

Corresponding author(s): Nima Aghaeepour

Last updated by author(s): Oct 22, 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No new data was collected for this study. No software was used to collect data for this study.
Data analysis	Code can be found at https://github.com/tpjoe/TPN2.0 The following packages and versions are relevant to our analysis: Python: 3.10, NumPy: 1.24.4, Pandas: 2.2.2, SciPy: 1.11.1, scikit-learn: 1.2.2, PyTorch: 2.0.1, TensorFlow: 2.11.0, statsmodels: 0.14.0, sksurv: 0.21, lifelines: 0.28.0, missForest: 3.1.3, forestplot: 0.4.1.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Stanford University and University of California, San Francisco are the custodians of their own individual patient-level protected information. Access to these data is subject to approval from the respective institutions. To request data access, researchers or institutions should contact N.A., Stanford University Privacy Office, and the Institutional Review Boards (IRBs). Requests will be reviewed on a rolling basis and timeline is dependent on the IRB. Data will be for research purposes only and must comply with ethical guidelines and institutional policies on data privacy.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Populations from Stanford and UCSF include both male and female sex. For both cohorts, their demographics are determined based on the available EHR data and are included as supplementary tables.
Reporting on race, ethnicity, or other socially relevant groupings	Race information was included in the manuscript as a part of the model's input and for post-hoc analysis.
Population characteristics	The details of the two populations (including age, sex, and race) are included as supplementary tables in the manuscript. For both cohorts, their demographics are determined based on the available EHR data.
Recruitment	Our study re-analyzes existing datasets, and no additional patients were recruited as part of our study.
Ethics oversight	The study was approved by Institutional Review Board of Stanford University (#39225) and of UCSF (#17-22929).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size calculation was performed prior to the study. Sample size was determined by the number of patients with available EHR data and fitted the inclusion criteria.
Data exclusions	Data of patients who received their first day of TPN after 2 years of age were excluded. We used 2 years as a cutoff (pre-established before the model development and analysis) as morbidities of interest mostly occur within this timeline.
Replication	The results were reproduced independently through an external validation at UCSF on their own data (tested once without cross validation or model retraining).
Randomization	Randomization was not performed as it is not possible in studies that utilize observational data such as ours. Participants were randomly allocated to training, testing, and validation sets.
Blinding	During external validation at UCSF, the researchers at Stanford were blinded and did not have access to the data.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks	N/A
Novel plant genotypes	N/A
Authentication	N/A