博士学位論文

Estimation of Threshold Dose Using a Change-Point Regression Model –Application for the Safety Assessment of Amino Acids by Systematic Review–

（変化点回帰モデルを用いた閾値量推定–システマティックレビューによるアミノ酸の安全性評価への応用–）

令和 5 年度

横浜薬科大学大学院薬学研究科

薬学専攻博士課程

倉持　優衣

CONTENTS

ABBREVIATIONS

ADI	acceptable daily intake
AE	adverse event
AIC	Akaike's information criterion
CPRM	change-point regression model
DHA	docosahexaenoic acid
EBM	evidence-based medicine
EPA	eicosapentaenoic acid
GRADE	Grading of Recommendations Assessment, Development and Evaluation
HOI	highest observed intake
L-Arg	L-Arginine
L-Cit	L-Citrulline
L-Lys	L-Lysine
LOAEL	lowest observed adverse effect level
MA	meta-analysis
NO	nitric oxide
NOAEL	no observed adverse effect level
OSL	observed safety level
omega-3 PUFA	omega-3 polyunsaturated fatty acid
RCT	randomized controlled trial
RD	risk difference
RoB	risk of bias
SF	safety factor
SR	systematic review
TD	threshold dose
UC	usual consumption
UL	upper limit
SF	safety factor
w-CPRM	weighted change-point regression model

Chapter 1.

INTRODUCTION

Amino acids are utilized for the biosynthesis of proteins, including structural proteins, metabolic enzymes, receptors, hormones and so on. They are also utilized for energy production and precursors for essential metabolites, such as nucleic acids, glutathione, several kinds of neurotransmitters, coenzyme A etc. Twenty different amino acids are employed in the body's protein biosynthesis process, categorized into nine essential amino acids and eleven nonessential ones. Since essential amino acids cannot be synthesized in the body, they have to be taken from foods. Thus, their nutritional requirements, i.e. Recommended Daily Allowance (RDA) for human have been set by the governments in the US, EU and other countries in addition to by WHO/FAO[1,2].

Although human consume each amino acid from foods every day, excessive intake of each amino acid could be harmful, as are in cases of other nutrients. Animal studies have shown that excessive intake of every single amino acid reduces appetite and causes growth retardation[3]. Adverse effects of excessive intake of single amino acids were also reported in human; bolus ingestion of for examples methionine 50 times higher amount than daily intake caused nausea, vomiting, and hepatic dysfunction, and finally death[4]. Thus, it is important to set Tolerable Upper Intake Levels (ULs) for individual amino acids in addition to their RDAs to reduce the risks of their deficiency and excess[1,2]. Nowadays necessity of ULs for amino acids seems to become bigger than ever, since individual single amino acids are now available in the market as supplements and since it is easy to ingest large amount of single amino acids much more than usual intake levels.

In general, Acceptable Daily Intakes (ADIs), i.e. safe intake levels, of chemicals in human is calculated from No Observed Adverse Effect Level (NOAEL) determined in experimental animals through various tests that enable the assessment of dose-response using default Safety Factor (SF) of 100, as follows:

$$ADI = SF \times NOAEL = 1/10 \times 1/10 \times NOAEL$$

However, it's important to note that this general approach to estimate ADI cannot be applied to determine safe intake levels of nutrients including amino acids. Table 1 presents the NOAEL from

rat studies, ADI calculated from the NOAEL, human requirement, mean intake in the US and mean intake in the top 10%) for each amino acid[5] (partly modified). Taking L-lysine (L-Lys) as an example, an essential amino acid, the ADI derived from the rat study is 34 mg/kg/day, which closely aligns with the human requirement of 38 mg/kg/day and much less than general intakes in the US. If the calculated ADI is compared to general intake levels in Japanese, the result is the same. Iwasaki et al. reported that the mean intake of L-Lys in Japanese individuals is approximately 5744 mg/day for men and 4789 mg/day for women[6], both are much higher than the calculated ADI (= 2024 mg/day in 60 kg body weight). Similar relations of the calculated ADIs with requirements and with general intake levels are seen in all the amino acids (Table 1). These results clearly indicate that ADIs or ULs of amino acids cannot be determined from animal data using the default SF and that the SF to extrapolate animal results to human should be less than 100. Then how much is the UF for amino acids?

In a previous study addressing the safety dosage of amino acids, Wu estimated the safety dose of L-arginine (L-Arg) using the FDA's conversion table. This involved determining the non-observed toxicity dose from the results of a non-clinical study, where L-Arg was administered to rats or pigs over a period of 90 days[7, 8]. The dose derived from the conversion table has been utilized as evidence for determining the initial amount administered to humans in what is known as a phase I study. Consequently, due to biological differences, it is insufficient to establish its safe upper intake limit, i.e. its UL. Blachier et al. approached the safety assessment of L-Arg by evaluating the NOAEL/UC ratio, utilizing two parameters: the NOAEL and usual consumption (UC)[5]. Unfortunately, many amino acids lack NOAEL/UC ratio data, rendering the results inconclusive. But a partial comparison of NOAEL/UC ratios among some amino acids has indicated that the safety range in humans is narrower than in animals. Based on these results, however, it is not sufficient to determine the upper safe limit simply from the conversion ratio. Thus, there is no scientific basis to generalize SF for amino acids.

Blachier et al. and Elango reviewed the NOAEL in humans[5, 9]. Safety assessments of amino acids are conducted in various ways but all of them are tentative. Although we believe that the evaluation of the data from human studies is important, few studies have been conducted in human to investigate effects of graded dosages of amino acids to assess their tolerable intake levels.

In recent years, systematic review (SR) has been used as an approach for assessing safety as well as effectiveness and usefulness. Hayamizu et al. used L-Lys as a model and considered the assessment of its safety by SR. They provisionally reported its NOAEL in healthy subjects as 6000 mg/day[10].

However, most of the studies that they identified have targeted humans, such as dose-finding trials, to reveal the nutritional requirements and the effect of L-Lys-supplemented fortified foods (wheat), since L-Lys is an essential amino acid. Meanwhile, they identified few intervention trials involving comparisons with a control group, such as Randomized Control Trials (RCT).

Comparison with a control group is needed for attempting to accurately characterize the occurrence of adverse events associated with an intervention. In addition, when performing a safety assessment by SR, it is possible to obtain the NOAEL by comprehensively collecting the results of different research. Therefore, the NOAEL we have reported is not the value estimated statistically approach, but we think that the value is nearly the same as the value called the observed safety level (OSL)[11].

We are currently conducting safety assessments for various amino acids, including L-ornithine, L-citrulline, and glycine, in addition to L-Lys. However, due to the limited availability of intervention studies, particularly those comparing target groups like RCT, we have chosen to use L-Arg as a model in our current research. To address the scarcity of intervention studies, we utilize a model available for estimating threshold doses (TDs)[12], considering the reported doses as the OSL.

Moreover, recognizing the heterogeneity present in each study—a common challenge in SR—we implemented a weighting approach to the model for estimating the TD.

The primary objective of our research is to develop a model that accounts for the heterogeneity between studies in the safety assessment of amino acids through systematic review. To achieve this, we initially apply the Change-Point Regression Model (CPRM), which allows for the estimation of the TD. Subsequently, we apply this model to estimate the TD of L-Arg in the safety assessment of amino acids.

Table 1 Comparing between ADI calculated from NOAEL in rat study and human requirement and mean intake.

	NOAEL in rat study (g/kg/day)	Calculated ADI (mg/kg/day)	Human Requirement (mg/kg/day)	Mean intake in the US (mg/kg/day)	Mean intake in the top 10% (mg/kg/day)	HOI (mg/kg/day)
Leu	3.33	33	42	105	135	500
IIe	1.57	16	19	61	78	225
Val	3.23	32	24	69	88	225
Met	0.24	2.4	19(Met+Cys)	31	40	63
Cys	0.50	5.0		17	22	-
Phe	1.55	15	33(Phe+Tyr)	59	75	152
Tyr	0.60	6.0		49	63	200
Lys	3.36	34	38	93	122	86
Thr	3.27	33	20	52	67	86
Trp	0.78	7.8	5	16	20	44
His	1.32	13	14	39	50	56
Ala	2.00	20	-	64	83	417
Asn	1.65	17	-	57	73	-
Asp	0.70	7.0	-	57	73	150
Arg	3.30	33	-	74	84	375
Gln	0.83	8.3	-	131	166	750
Glu	-	-	-	131	166	1600
Ser	2.77	28	-	60	76	400
Gly	2.00	20	-	68	89	800
Pro	2.77	28	-	89	108	488

NOAEL: no observed adverse effect level, ADI: acceptable daily intake,
HOI: highest observed intake

REFERENCES

1. FAO/WHO/UN. 2007. Protein and amino acid requirements in human nutrition: report of a joint FAO/WHO/UNU expert consultation. Geneva, Switzerland: WHO Press.
2. Food and Nutrition Board, Institute of Medicine, National Academies. 2005. Dietary Reference Intakes for Energy, Carbohydrate, Fiber, Fat, Fatty Acids, Cholesterol, Protein, and Amino Acids. Washington, D.C.: The National Academy Press.
3. Anderson SA and Raiten DJ, 1992. Safety of Amino Acids Used as Dietary Supplements. Prepared for Food and Drug Administration by Life Sciences Research Office. Federation of American Societies for experimental biology, Bethesda, MD, USA.
4. Garlick PJ. The nature of human hazards associated with excessive intake of amino acids. J Nutr. 2004 Jun;134(6 Suppl):1633S-1639S; discussion 1664S-1666S, 1667S-1672S.
5. Blachier F, Blais A, Elango R, Saito K, Shimomura Y, Kadowaki M, Matsumoto H. tolerable amounts of amino acids for human supplementation: summary and lessons from published peer-reviewed studies. Amino Acids 2021; 53: 1313-1328
6. Iwasaki M, Ishihara J, Takachi R, Todoriki H, Yamamoto H, Miyano H, Yamaji T, Tsugane S. Validity of a self-administered food-frequency questionnaire for assessing amino acid intake in Japan. J Epidemiol 2016; 26:36-44
7. Wu Z, Hou Y, Hu S, Bazer FW, Meininger CJ, McNeal CJ, Wu G. Catabolism and safety of supplemental L-Arg in animals. Amino Acids 2016; 48: 1541-1552
8. Guidance for Industry Estimating the Maximum Safe Starting Dose in Initial Clinical Trials for Therapeutics in Adult Healthy Volunteers (2005) https://www.fda.gov/media/72309/download Accessed 9 May 2023
9. Elango R. Tolerable Upper Intake Level for Individual Amino Acids in Humans: A Narrative Review of Recent Clinical Studies. Adv Nutr. 2023; 14 :885-894.
10. Hayamizu K, Oshima I, Fukuda Z, Kuramochi Y, Nagai Y, Izumo N, Nakano M. Safety assessment of L-lysine oral intake: a systematic review. Amino Acids 2019; 51 :647-659
11. Shao A, Hathcock JN. Risk assessment for the amino acids taurine, L-glutamine and L-Arg. Regul Toxicol Pharmacol 2008; 50 :376-99
12. Lutz WK, Lutz RW. Statistical model to estimate a threshold dose and its confidence limits for the analysis of sublinear dose-response relationships, exemplified for mutagenicity data. Mutat Res 2009; 678(2):118-22.

Chapter 2.

A change-point regression approach for estimating the threshold dose from a systematic review

2.1 BACKGROUND

Systematic review (SR) is a method of systematically collecting, organizing, summarizing, and integrating past independent studies to estimate the effects of interventions and the risks of exposure[1]. With the development of evidence-based medicine (EBM), the results of SRs have been given more weight than independent single research reports[2]. Additionally, the results of SRs can be used to determine whether there is a need to start new research, enabling the avoidance of unnecessary studies[3].

SRs on not only pharmaceuticals but also food ingredients have been reported, such as the effects of chromium supplementation for glucose metabolism in diabetic patients[4], the role of vitamin D in COVID-19 patients[5], and the effects of vitamins C and E on muscle pain in healthy adults[6].

SRs can also be used to evaluate safety. In the case of pharmaceuticals, safety is often reported along with efficacy and usefulness. However, few reports on SRs on the safety of food ingredients have been published. To assess the safety of food constituents, it is important to estimate the upper limit (UL) at which we can use these constituents safely, that is, the no observed adverse effect level (NOAEL) and the lowest observed adverse effect level (LOAEL).

Hayamizu *et al.*[7] proposed a method to estimate the NOAEL using SR and reported the NOAEL of L-lysine, one of the amino acids, as a model. They determined the NOAEL and LOAEL as observed values based on information on the incidence of adverse events and the dosage from each study included in SR. However, the estimated NOAEL is substantially equivalent to a value called the observed safety level (OSL)[8] rather than a statistically estimated value, therefor it remained issue for determining the safety use level.

Here, we consider estimation of the maximum dose of no adverse events occur by SR.

The dose dependence of adverse events in a meta-analysis (MA) can be confirmed by meta-regression analysis[9]. This does not involve a simple regression model, but is rather a method of analysis that incorporates the heterogeneity among studies into the model. In MA, a fixed-effects model that assumes that random errors are the difference between studies or a random-effects model that adds non-negligible heterogeneity of the difference to the above model is generally used. We think that it is possible to consider the dose response in a manner more strongly based

on actual conditions by using a random-effects model. The formula of the model is shown below:

$$\hat{\theta}_k = \theta + \beta_1 \chi_{1k} + \cdots + \beta_P \chi_{Pk} + u_k + \sigma_k \epsilon_k, \quad \epsilon_k \sim N(0,1); u_k \sim N(0,\tau^2)$$

with *K=1, ..., K* and independent error terms *u* and ε [10, 11]. As this model includes both fixed-effects (βs) and random-effects terms (u_k with variance τ^2) this meta-regression model is also called a mixed effects model. Fixed-effects meta-regression is a special case of a mixed-effects model when the between-study variance $\tau^2 = 0$.

Besides meta-regression analysis, subgroup analysis might also be used for confirming the dose dependence. Subgroup analyses aim to assess whether effects are similar among specific patient groups or whether changes occur in specific patient characteristics[12]. Dose response by subgroup analysis can be confirmed by the forest plot, which is based on the results of classification and integration depending on the dose used in each study. However, the primary purpose of safety studies is to find the threshold at which toxicity starts to occur[13]. Meta-regression analysis and subgroup analysis can observe the dose response but cannot directly determine the dose threshold beyond which adverse events occur (Fig. 1).

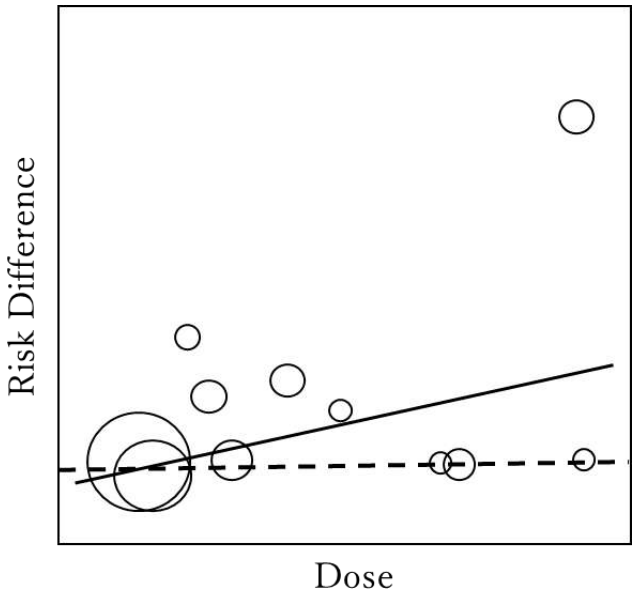

Fig. 1 An example of meta-regression analysis.
This panel can be used to observe the dose response, but not directly determine the threshold beyond which adverse events occur. Bubble size shows 1/SE.

In the field of toxicology, a linear model and a quadratic model have been used as response models to identify such a threshold, but a hockey stick model is now a particularly representative

option[14]. The formula of a hockey stick model is shown below, with P representing the probability of event occurs.

$$P = \begin{cases} 0, & for\ \chi \ll \chi_{cp} \\ F_\theta[\alpha + \beta log(\chi - \chi_{cp})] & for\ \chi > \chi_{cp} \end{cases}$$

x is the dosage and x_{cp} is the dosage of the change point. The probability of reaction (P) is 0 as long as the dose is the change point or under. However, it is $\alpha+\beta\log(x-x_{cp})$ if the dose exceeds the change point[13, 15].

The model is characterized by a biphasic manner, with regions that are not dose-responsive and regions that are dose-responsive. Theoretically, the threshold dose (TD) should be between NOAEL and LOAEL (Fig. 2).

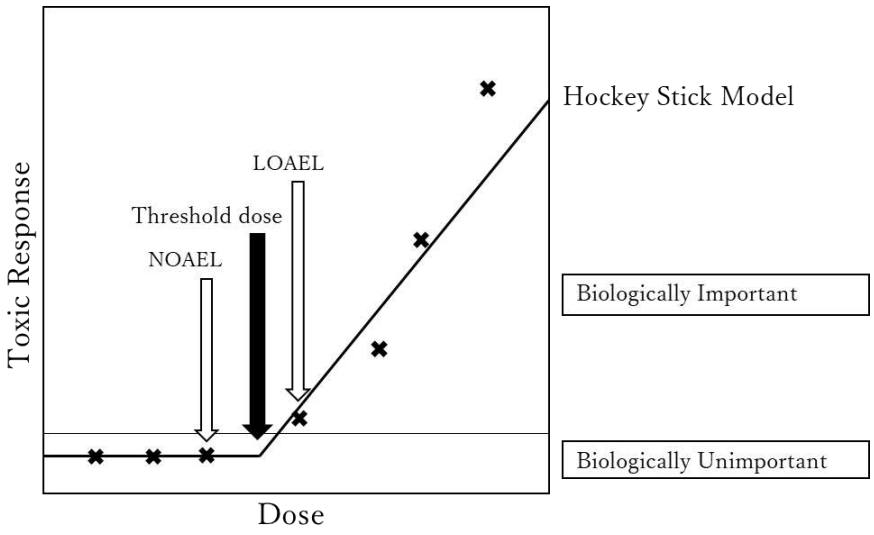

Fig. 2 An example of the hockey stick model. The model is characterized by a biphasic manner, with regions that are not dose-responsive and regions that are dose-responsive. Theoretically, the threshold dose (TD) should be between NOAEL and LOAEL.

The threshold can be determined from the point at which these two straight lines meet. As an example of application of the hockey stick model, it has been used to determine the mercury dose abnormality in children and the amount of exposure in the fetal period using the concentration of mercury in the mother's hair[16].

In recent years, the change-point regression model (CPRM) has been proposed as a method of analysis to lead the threshold of dose response. The CPRM has been used for analyzing clinical trials, such as evaluating the change of visceral fat area upon taking (-)-HCA[17], estimating the

required amount of L-lysine[18], estimating the inter-individual variation of the protein requirement by the indicator amino acid oxidation method[19], and estimating the amino acid requirement adjusted with a carryover effect[20]. In these studies, identification of the dose threshold is regarded as important. Applying the CPRM to the hockey stick model, the following equation can be used:

$$y_i = \beta \mathrm{I}(\chi_i > \chi_{cp})(\chi_i - \chi_{cp})$$

y_i is the probability of reaction and x_i is the dosage. $I(\ \cdot\ ) = 1$ if $x_i > x_{cp}$ and is 0 otherwise.

Therefore, it is considered to be possible to estimate these thresholds by CPRM even in searching for the threshold beyond which adverse events occur.

Incidentally, in clinical study, research results from RCTs are highly reliable, and SR reports, which integrate such research results, are interpreted to have the highest evidence level. However, each clinical trial used for SRs often has internal validity, featuring a mixture of low and high quality in terms of the study design and data handling. These features are represented by assessing the risk of bias (RoB) in RCTs, as an index of study quality.

In our previous studies, we conducted RoB assessment using the estimation method recommended by the Cochrane Collaboration[21] or the Jadad score[22]. We confirmed that there is large variability in the RoB of studies included in a meta-analysis irrespective of which of these assessments is performed. Accordingly, upon estimating the TD by SR, we also need to consider RoB-based weighting as a heterogeneous feature of each included study. However, to the best of our knowledge, no reports have been published about practical examples of CPRM performed by assigning weights to each set of data.

In the present study, we propose CPRM with weighting by the heterogeneity of the studies to assess the safety of food components by SR. We also report the results of applying the weighted CPRM to the TD estimation using gastrointestinal symptoms as an indicator from an SR study of omega-3 fatty acid intake.

2.2 MATERIALS AND METHODS

For finding out the value that considered with heterogeneity of each study, we propose the following regression model, weighted CPRM (w-CPRM), which takes into account differences in study quality by using the method of weighted least squares:

$$\omega_i y_i = \omega_i \beta \mathrm{I}(\chi_i > \chi_{cp})(\chi_i - \chi_{cp}) + \omega_i \varepsilon_i$$

where y_i and x_i are the risk difference and dose for the i-th study, respectively, and x_{cp} is the unknown change point, or threshold. $I(\cdot)$ is an indicator function defined as $\mathrm{I}(x_i > x_{cp}) = 1$ if $x_i > x_{cp}$ and is 0 otherwise. ω_i is the weight of the i-th study. Random error ε_i follows a normal distribution with mean 0 and variance σ^2; it is assumed to be statistically independent.

Here, the model is indicated as:

$$\omega_{1i} y_i = \omega_{1i} \beta \mathrm{I}(\chi_i > \chi_{cp})(\chi_i - \chi_{cp}) + \omega_{1i} \varepsilon_i \qquad (Eq.1)$$

if the weight of the i-th study on the fixed-effects model is ω_{1i}. It is evaluated based on data in the format of a 2×2 table, when the incidence of side effects is the outcome [11].

This 2×2 table is shown below:

	Outcome1	Outcome2	
Group1	a_i	b_i	n_{1i}
Group2	c_i	d_i	n_{2i}

where a_i, b_i, c_i, and d_i denote the cell frequencies and n_{1i} and n_{2i} the row totals in the i-th study. The risk difference is equal to (a_i/n_{1i})- (c_i/n_{2i}).

From the above, the weight ω_{1i} of each study on the fixed-effects model is calculated below. The weight of each study is proportional to the total number of subjects.

$$\omega_{1i} = \frac{1}{SE_i^2}$$

$$SE_i = \sqrt{\frac{a_i b_i}{n_{1i}^3} + \frac{c_i d_i}{n_{2i}^3}}$$

In addition, the model using ω_{2i} is described:

$$\omega_{2i}y_i = \omega_{2i}\beta \mathrm{I}(\chi_i > \chi_{cp})(\chi_i - \chi_{cp}) + \omega_{2i}\varepsilon_i \qquad (Eq.2)$$

Then, ω_{2i} is:

$$\omega_{2i} = \omega_{1i} + RS_i = \frac{1}{SE_i^2} + RS_i$$

$$RS_i = \frac{score_i}{(\sum_{i=1}^{n} score_i)}$$

RS_i is derived using the results of the RoB assessment for each trial. The RoB evaluation by the Cochrane Collaboration is rated on three levels, low, high, and unknown, for each bias risk item. We scored these results as "low risk" = 1, "high risk" = –1, and unclear = 0 and set $score_i$ as the value upon summing them up. RS_i was calculated as the ratio of $score_i$ to the total $score_i$ of all included studies. ω_{2i} is the value that added ω_{1i} to RS_i that calculated from the result of RoB. The result of RoB assessment is considered to one of heterogeneities, and if it is, we need to add them to weights. ω_{2i} was calculated by adding RS_i to ω_{1i}, resulting in a total of 200%.

In CPRM that considers the heterogeneity of between studies, each model was compared by the maximum likelihood method used Akaike's information criterion (AIC) as an indicator. The formula of AIC is:

$$\mathrm{AIC} = -2\text{maximum log-likelihood} + 2p$$

where p is the number of unknown parameters of the model[23]. x_{cp} with minimum AIC can be determined to be the value closest to the TD. In this study, we also compared the CPRM with no adjustment by weighting and the CPRM weighted by ω_{1i} or ω_{2i}.

2.2.1 Data synthesis and statistics

Statistical analysis was conducted by R4.1.4 (www.r-project.org) and using the packages meta and metafor[10,11].

2.2.2 Application data (omega-3 study)

Omega-3 polyunsaturated fatty acid (omega-3 PUFA) is one of the essential nutrients for maintaining health. As main types of omega-3 PUFAs, α-linolenic acid, which is contained in plant-based food, and eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA), which are

contained in fish, were selected here. The risk of cardiovascular diseases such as heart disease and stroke are considered to be decreased by consuming fish or taking omega-3 supplements.

In this study, we used reported data on the side effects from "Omega-3 fatty acids for the primary and secondary prevention of cardiovascular disease (review)" reported by Abdelhamid *et al.*[24] as extrapolation data.

We usually have focused on the safety of the food components such as supplement. This review had evaluated the safety not only the effectiveness. The risk bias of each study we need to insert to the model have already evaluated. All of study were RCTs and there was some difference score between each study. Also, the number of target study is not less so we judged it appropriate this study to apply.

The RoB evaluation by the Cochrane Collaboration is rated on three levels, low, high, and unknown, for each bias risk item. We scored these results as "low risk" = 1, "high risk" = −1, and unclear = 0 and set $score_i$ as the value upon summing them up. RS_i was calculated as the ratio of $score_i$ to the total $score_i$ of all included studies. ω_{2i} was calculated by adding RS_i to ω_{1i}, resulting in a total of 200%.

In CPRM that considers the heterogeneity of each study, each model was compared by the maximum likelihood method used Akaike's information criterion (AIC) as an indicator. The formula of AIC is:

We usually have focused on food and supplement without medicine and the review has already estimated the data of risk bias we need to insert to CPRM and all of them were RCT in the study design.

This SR targeted adult males and females with a high risk of cardiovascular disease, with the exclusion of patients with acute diseases and pregnant women. In terms of their design, all included studies were randomized controlled trials (RCTs). Side effects are shown as tertiary outcomes in this review. The integrated data compare two groups with high and low omega-3 fatty acid intake. However, in some studies, the low-intake group was given intensive advice about the intake of omega-3 fatty acid by eating fish, instead of a direct intervention involving supplements or a placebo. For that reason, these studies were excluded. Here, we used 24 sets of data (excluding the above) from 29 studies reporting the gastrointestinal side effects of omega-3 intake (number of papers: 28). The dose of omega-3 fatty acid was 376–5500 mg/day and the dosing period was 12–72 months (Table 1). In this research, we considered the application of w-CPRM by using data from studies on the incidence of digestive symptoms as a side effect of omega-3 intake.

2.3 RESULTS

In their study on omega-3, Abdelhamid et al. presented the risk of occurrence of digestive symptoms as a risk ratio. However, we instead calculated the risk difference (RD), since this avoids the problem of zero cells. We also added 0.0001 to all values to reflect the influence of the weight by ω_{1i} and ω_2 even in cases where RD was 0. We assumed that the incidence of adverse events until reaching the threshold of the expression is 0 and modified the intercept of w-CPRM so that it goes through 0.

The results of the RoB assessment used for weighting correction were those already reported by the Cochrane Collaboration[21]. The RoB consists of nine items, as follows: random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessors, use of incomplete outcome data, selective outcome reporting, attention bias, limited compliance, and other bias. Each item in the RoB is rated on three levels: "low risk", "high risk", and "unclear" that can't assort either high and low risk. We scored each category as low risk = 1, high risk = −1, and unclear = 0 and obtained the overall score by summing them (maximum value: 9). We considered this value as $score_i$ and the range was 3–8. It shows that there was heterogeneity in the quality of the studies even though all studies were RCT (Tables 1 and 2).

Here, the formula of unweighted CPRM is as follows:

$$y_i = \beta \mathrm{I}(\chi_i > \chi_{cp})(\chi_i - \chi_{cp}) + \varepsilon_i \quad \text{(Eq. 3)}$$

The results of CPRM for Omega-3 study analyzed in each model are shown in Fig. 3.

Panels A–C indicate the results of Eqs. 1–3, respectively. The size of the circle indicates ω_1 and ω_2 of each study. Panel C is drawn by an unweighted model, so the sizes of the circles are the same.

Note that, for toxicities such as carcinogenicity, there is sometimes no threshold. Therefore, models weighted by ω_{1i} and ω_{2i} with the assumption of there being no threshold were also established, as shown below, respectively:

$$\omega_{1i} y_{1i} = \omega_{1i}\alpha + \omega_{1i}\beta\chi_{1i} + \omega_{1i}\varepsilon_i \quad \text{(Eq. 4)}$$

$$\omega_{2i} y_{2i} = \omega_{2i}\alpha + \omega_{2i}\beta\chi_{2i} + \omega_{2i}\varepsilon_i \quad \text{(Eq. 5)}$$

The results of regression model obtained by Eqs. 4 and 5 are shown in Fig. 4. Panels A and B indicate the results of Eqs. 4 and 5, respectively. The size of the circle indicates ω_1 and ω_2 of each

study. To evaluate the usefulness of w-CPRM, we compared Eqs. 1 and 4 using ω_{1i} and Eqs. 2 and 5 using ω_{2i}, and examined whether there was any threshold regarding the dose above which adverse events begin to occur (Table 3). The AIC values of Eqs. 1, 2, 4, and 5 are −100.79, −64.99, −93.94, and −61.29, respectively. Compared with Eqs. 4 and 5, which are linear regressions, Eqs. 1 and 2, which are w-CPRM regressions, showed improvements in AIC. This suggested that there is a threshold for the occurrence of gastrointestinal symptoms upon the intake of omega-3 fatty acids as evaluated in this study, regardless of the weighting.

The minimum AIC of Eq. 3 (unweighted CPRM) was −58.14. We confirmed that the AIC of both Eq. 1 and Eq. 2 using weighting fitted better than that of Eq. 3. The x_{cp} values were 2752, 2313, and 2499 mg/day, respectively, which were estimated to be the TD in each model. Confirming an AIC profile, the minimum value of x_{cp} observed between 2000 and 3000 mg in any case (Fig. 5). However, the bottom of Eq. 1 profile was wider and the line close to the minimum value goes up gentler, the estimated range of the TD with Eq. 1 seemed to be wide than those for Eqs. 2 and 3. (Fig. 5).

Table 1 Summary of identified studies for gastro intestine symptoms in Omega-3 study

Study	higher Omega-3 (n)		lower Omega-3 (n)		dose (mg/day)	duration (month)	RD	ω_{1i}	$score_i$	RS_i	ω_{2i}
	GI sympton(+)	total	GI sympton(+)	total							
ADCS_2010	18	238	10	167	1020	18	0.0201	0.9	8	5.41	6.31
AFFORD_2013	6	153	5	163	2400	12	0.0101	1.3	3	2.03	3.33
AlphaOmega-EPA_DHA_2010	18	1192	10	1236	376	40	0.0101	13.7	7	4.73	18.43
AREDS2_2014	119	2147	145	2056	1000	60	-0.0201	7.3	8	5.41	12.71
EPE-A_2014	67	168	40	75	1800	12	-0.1301	0.1	6	4.05	4.15
EPIC-1_2008	45	187	41	184	3000	52	0.0201	0.3	5	3.38	3.68
EPIC-2_2008	65	189	62	188	3000	52	0.0101	0.2	5	3.38	3.58
FORWARD_2013	6	289	8	297	850	12	-0.0101	3.2	6	4.05	7.25
FOSTAR_2016	67	101	62	101	4500	24	0.0501	0.1	6	4.05	4.15
GISSI-HF_2008	96	3494	92	3481	866	46.8	0.0001	15	6	4.05	19.05

GI: Gasto intestine, RD:Risk difference

Table 1 Summary of identified studies for gastro intestine symptoms in Omega-3 study (continued)

Study	higher Omega 3 (n)		lower Omega 3 (n)		dose (mg/day)	duration (month)	RD	ω_{1i}	$score_i$	RS_i	ω_{2i}
	GI sympton(+)	total	GI sympton(+)	total							
Lorenz-Meyer_1996	10	70	2	63	5100	12	0.1101	0.3	5	3.38	3.68
MAPT_2017	175	820	164	832	1025	36	0.0201	1.4	8	5.41	6.81
OPAL_2010	0	434	4	433	700	12	-0.0101	11.7	7	4.73	16.43
ORIGIN_2012	14	6281	24	6255	840	72	0.0001	23.3	8	5.41	28.71
ORL_2013	27	171	24	165	3400	12	0.0101	0.4	5	3.38	3.78
Proudman_2015	1	86	1	53	5500	12	0.0501	0.1	6	4.05	4.15
Puri_2005	17	67	14	68	1900	12	0.0501	0.1	5	3.38	3.48
Raitt_2005	11	100	12	100	1300	24	-0.0101	0.3	5	3.38	3.68
Risk & Prevention_2013	200	6239	186	6266	870	60	0.0001	17.4	4	2.70	20.10
Rossing_1996	3	18	1	18	4600	12	0.1101	0.1	6	4.05	4.15

GI: Gasto intestine, RD:Risk difference

Table 1 Summary of identified studies for gastro intestine symptoms in Omega-3 study (continued)

Study	higher Omega 3 (n)		lower Omega 3 (n)		dose (mg/day)	duration (month)	RD	ω_{1i}	$score_i$	RS_i	ω_{2i}
	GI sympton(+)	total	GI sympton(+)	total							
SCIMO_1999	4	112	3	111	2000	24	0.0101	1	7	4.73	5.73
Shinto_2014	0	13	3	13	1650	12	-0.2301	0	8	5.41	5.41
SOFA_2006	17	273	12	273	800	12	0.0201	1.5	8	5.41	6.91
Tande_2016	18	64	28	63	2000	12	-0.1601	0.1	6	4.05	4.15

GI: Gasto intestine, RD:Risk difference

Table 2 The scoring items for calculating of $score_i$

Bias※	Authors' judgement		
	low risk	Unclear	high risk
1.Random sequence generation (selection bias)	+1	0	-1
2.Allocation concealment (selection bias)	+1	0	-1
3.Blinding of participants and personnel (performance bias)	+1	0	-1
4.Blinding of outcome assessment (detection bias)	+1	0	-1
5.Incomplete outcome data (attrition bias)	+1	0	-1
6.Selective reporting (reporting bias)	+1	0	-1
7.Attention	+1	0	-1
8.Compliance	+1	0	-1
9.Other bias	+1	0	-1

※The bias items were referred from omega-3 study

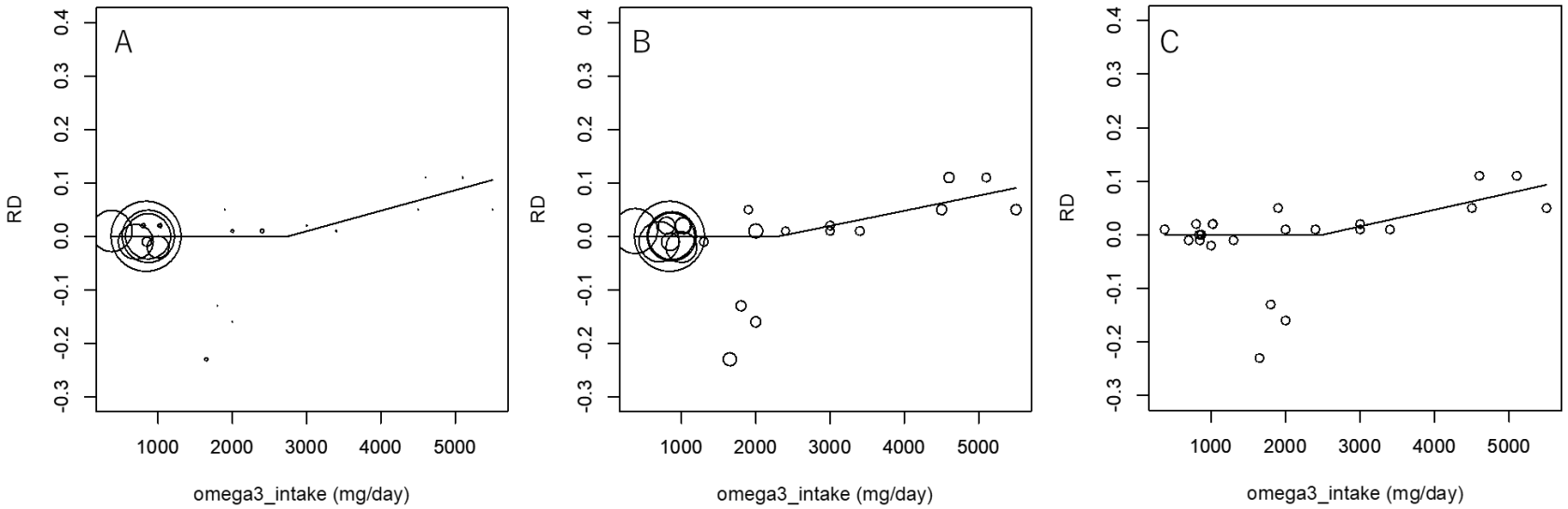

Fig. 3 Results of change-point regression model (CPRM) for omega-3 study. Panels A–C indicate the results of Eqs. 1–3, respectively. The size of the circle indicates ω_1 and ω_2 of each study. Panel C is drawn by an unweighted model, so the sizes of the circles are the same. RD: risk difference.

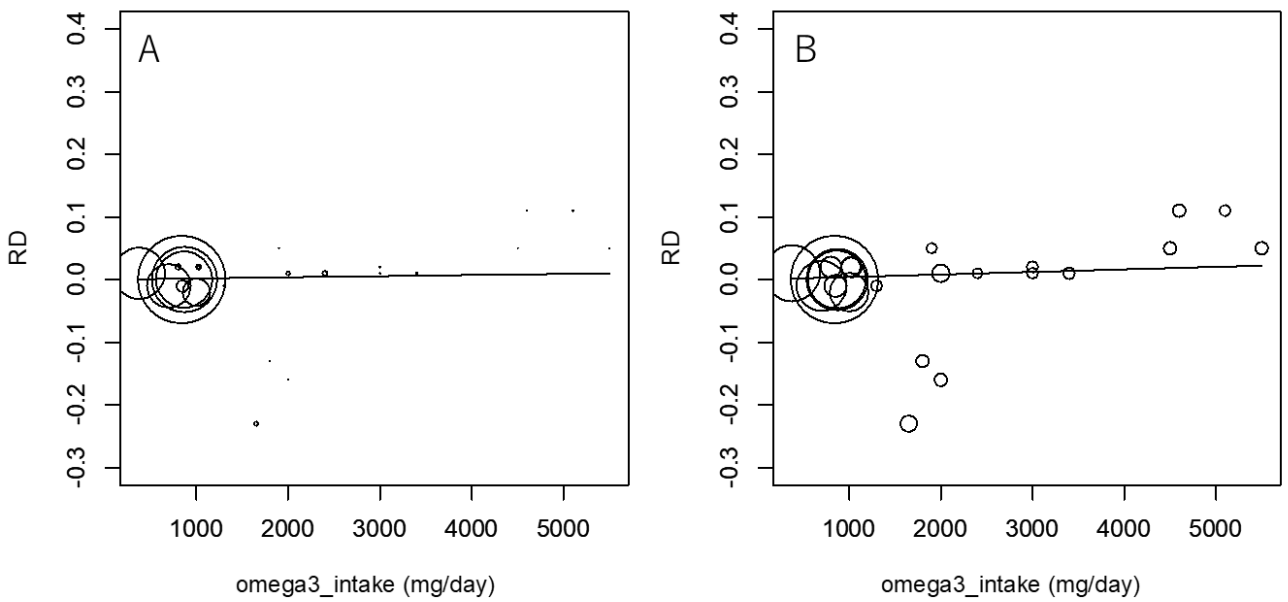

Fig. 4 Results of regression model for omega-3 study. Panels A and B indicate the results of Eqs. 4 and 5, respectively. The size of the circle indicates ω_1 and ω_2 of each study. RD: risk difference.

Table 3 Summary of result

model	Weight	x_{cp}	β	p-value	AIC
Eq. 1	ω_1	2752	3.84E-05	0.0091	-100.79
Eq. 2	ω_2	2313	2.84E-05	0.0483	-64.99
Eq. 3	-	2499	3.11E-05	0.0297	-58.14
Eq. 4	ω_1	-	1.83E-06	0.5250	-93.94
Eq. 5	ω_2	-	4.05E-06	0.5110	-61.29

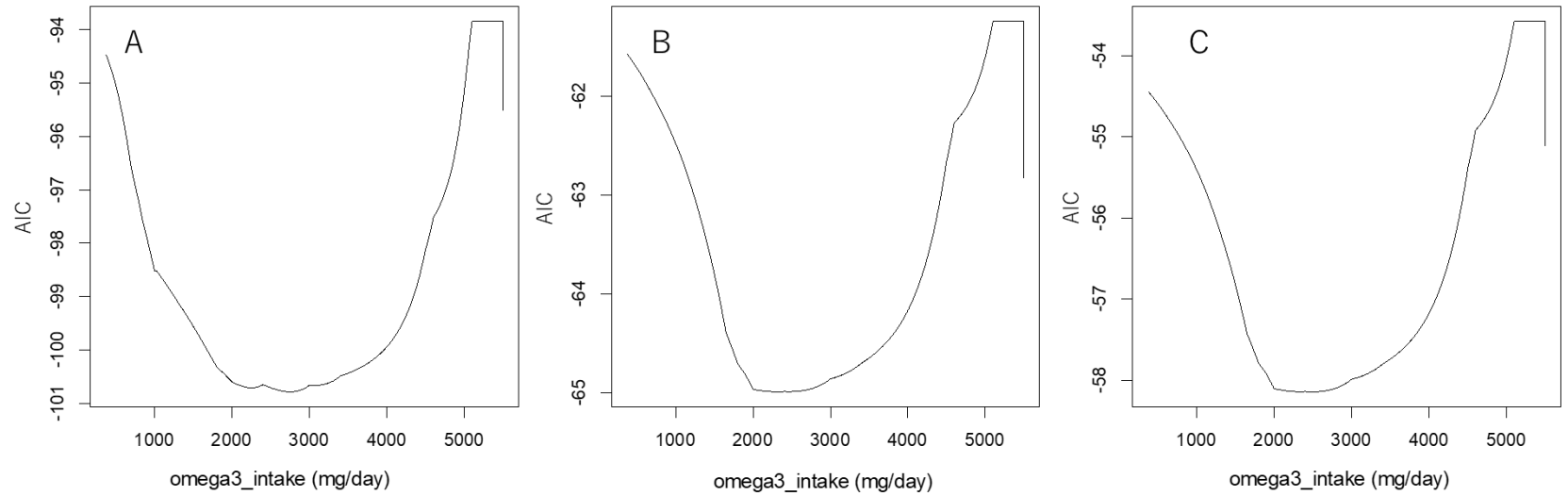

Fig. 5 AIC profiles of Eqs. 1–3. Panels A–C represent Eqs. 1–3, respectively. The change points of Eqs. 1–3 are estimated from minimum AIC as 2752, 2313, and 2499 mg/day, respectively.

2.4 DISCUSSION

In this study, we considered the application of w-CPRM in efforts to determine the TD based on SR. As an example, using the data from a study on omega-3, we used the result of RoB evaluated by the Cochrane Collaboration to obtain a weighted score. All of the clinical studies on omega-3 were RCTs, but $score_i$ used as an index of the RoB varied from 3 to 8 among these studies. This indicated the presence of variability in the quality of each study, even exclusively among clinical studies in the form of RCTs. Therefore, for the use of SRs in a wider field, it is considered necessary to evaluate the difference in quality of each study regardless of the study design. When there is large variability in the quality of studies included in a meta-analysis, we can consider performing subgroup analysis with the subgroups allocated based on the quality of the research. However, there is a problem that the number of studies is reduced by dividing them into subgroups. In addition, it is difficult to find the TD via this approach. Therefore, instead of conducting subgroup analysis, we proposed a method of estimating the TD by scoring the differences in the quality of the studies as weights and incorporating them into the CPRM, which allows the exploration of the TD.

In this study, ω_{1i} and ω_{2i} are used as the weights. ω_{1i} is the same as the weighting in the fixed-effects model. The weight for each study in the random-effects model is as follows.

$$\text{Weight}_i = \frac{1}{SE_i^2 + \widehat{\tau^2}}$$

The size of the variability of each study τ^2 is determined by the DerSimonian-Laird method[25, 26]. However, τ^2 is also calculated the same as SE_i by only incidence of the risk outcome and the total number of subjects; it does not include the RoB result, which evaluates the quality of the study. From the results of the present $score_i$, we confirm that there is large variation in the quality of the studies.

Interpreting this as inter-study error, in the omega-3 study, the results of the RoB evaluation are taken as RS_i and incorporated into ω_{2i} instead of τ^2. Since the larger the value of RS_i, the higher the quality of the test and the larger the weighting, it is added to ω_{1i}. In this way, the results of the RoB evaluation are reflected in the weighting as the heterogeneity of each test.

In this study, we compared models based on Eqs. 1–5 using the AIC, depending on the presence or absence of a threshold value and the weighting (Table 3). First, we determined the presence or absence of a threshold regarding the dose above which adverse events begin to occur by comparing the models with the same weighting. Here, Eqs. 4 and 5 are regression models using the weighted least squares method, and Eq. 4 in particular is consistent with meta-regression

analysis[27].

From comparison of the minimum AIC of each model, if the weighting is the same type (ω_{1i} or ω_{2i}), CPRM (Eqs. 1 and 2) clearly achieves better fitting than the linear regression model (Eqs. 4 and 5). We thus considered CPRM with a threshold to be suitable for analyzing the adverse events associated with the intake of omega-3 fatty acids.

This comparison process is required as the first step in the algorithm searching for the TD from SR. Although a smaller value of AIC indicates a better fit of the model, there is no consensus about the standard value of AIC at which a model is judged to have been improved. Therefore, comparisons with different models using the same weighting type (e.g., Eq. 1 vs. Eq. 4) are appropriate.

However, in comparisons with the same model using different weighting types (e.g., Eq. 1 vs. Eq. 2), it is inappropriate to make judgments directly based on AIC. On the other hand, comparison of Eqs. 1, 2, and 3 shows that x_{cp} is changed by weighting. This indicates that the change point may be significantly moved by weighting. The AIC profile (Fig. 5) shows the lowest value within a similar range in all models. This suggests the necessity of considering a more appropriate weighting method for a more detailed search of the change point.

Incorporating the results of RoB evaluation into w-CPRM should enable more realistic TD determination. In the procedure of SR, RoB is one of the steps requiring the most time and effort. However, we had not used inserting to something such as the model directly to estimate TD by just confirming its heterogeneity of studies included in a meta-analysis. The information on the incidence of adverse events and doses reported from studies with a high $score_i$ is reliable. Therefore, we believe that using the results of the RoB assessment as weights in the analysis would lead to more accurate conclusions. We could also consider that the result corrected by ω_{2i} needs to be used in the estimation of the TD in the future. In addition, we think that w-CPRM can be used in various research fields by devising values other than ω_{1i} and ω_{2i} used as weighting. To date, we have also conducted SRs in a variety of research fields, involving a number of heterogeneities such as the customs of the participants and their cultures, not just the use of RoB as incorporated into this study. Therefore, the adjustment of background factors as confounders could also potentially be applied.

To our knowledge, this is the first report describing the implementation of w-CPRM adjusted by weighting. Therefore, the validity of the values incorporated as weights and the method for judging the appropriateness of the weighting requires further investigation. In addition, the method that takes into account the quality of the study may be applied not only to the field of food and dietary supplements but also to a wide range of other research fields.

2.5 REFERENCES

1 Ng KH, and Peh WC. Writing a systematic review. *Singapore Med J* 2010; 51: 362-366

2 Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ* 1996; 312: 71-72

3 Swedish Council on Health Technology Assessment. Assessment of Methods in Health Care: A Handbook [Internet]. Stockholm: Swedish Council on Health Technology Assessment (SBU); 2016. SBU Method No. 0010.

4 Suksomboon N, Poolsup N, Yuwanakorn A. Systematic review and meta-analysis of the efficacy and safety of chromium supplementation in diabetes. *J Clin Pharm Ther* 2014; 39: 292-306.

5 Rawat D, Roy A, Maitra S, Shankar V, Khanna P, Baidya DK. "Vitamin D supplementation and COVID-19 treatment: A systematic review and meta-analysis". *Diabetes Metab Syndr* 2021; 15: 102189.

6 Torre MF, Martinez-Ferran M, Vallecillo N, Jiménez SL, Romero-Morales C, Pareja-Galeano H. Supplementation with Vitamins C and E and Exercise-Induced Delayed-Onset Muscle Soreness: A Systematic Review. *Antioxidants (Basel)* 2021; 10: 279

7 Hayamizu K, Oshima I, Fukuda Z, *et al*. Safety assessment of L-lysine oral intake: a systematic review. *Amino Acids* 2019; 51: 647-659

8 Hathcock JN, and Shao A. Expanded approach to tolerable upper intake guidelines for nutrients and bioactive substances. *J Nutr* 2008; 138: 1992S-1995S

9 Thompson SG, and Higgins JP. How should meta-regression analyses be undertaken and interpreted? *Stat Med* 2002; 21: 1559-1573

10 Schwarzer G, Carpenter JR, Rücker G. *Meta-Analysis with R.* Switzerland: Springer International Publishing, 2015; 85-104

11 Viechtbauer, and Wolfgang. Conducting Meta-Analyses in R with the metafor Package. *Journal of Statistical Software* 2010; 36: 3.

12 Thabane L, Mbuagbaw L, Zhang S, *et al*. A tutorial on sensitivity analyses in clinical trials: the what, why, when and how. *BMC Med Res Methodol* 2013; 13: 92

13 Cox C. Threshold dose-response models in toxicology. *Biometrics* 1987; 43: 511-523

14 Johnson GE, Doak SH, Griffiths SM, *et al*. Non-linear dose-response of DNA-reactive genotoxins: recommendations for data analysis. *Mutat Res* 2009; 678: 95-100

15 Yanagimoto T, and Yamamoto E. Estimation of safe doses: critical review of the hockey stick regression method. *Environ Health Perspect* 1979; 32: 193-199

16 Grandjean P and Budtz-Jørgensen E. Epidemiological Approaches to Metal Toxicology. In: Nordberg GF, Fowler BA, Nordberg M, eds. *Handbook on the Toxicology of Metals Fourth Edition Volume I: General Considerations*, Amsterdam: ELSEVIER, 2014; 265-279

17 Hayamizu K, Yamashita N, Hattori S, Kakuma T. A change-point regression approach for efficacy evaluation of dietary supplements. *J Clin Biochem Nutr* 2009; 44: 285-290

18 Hayamizu K, Kato M, Hattori S. Determining amino acid requirements from repeated observations on indicator amino acid oxidation method by mixed-effect change-point regression models. *J Clin Biochem Nutr* 2011; 49: 115-120

19 Hayamizu K, Aoki Y, Izumo N, Nakano M. Estimation of inter-individual variability of protein requirement by indicator amino acid oxidation method. *J Clin Biochem Nutr* 2021; 68: 32-36

20 Kato M, Hattori S, Hayamizu K. Estimation of Amino Acid Requirement Adjusting for Carry-Over Effect Based on Approximate Change-Point Regression Model. *Biomedical Research* 2013; 24: 216-222

21 Higgins JP and Altman DG. Assessing risk of bias in included studies. In: Higgins JPT and Green S, eds. *Cochrane Handbook for systematic reviews of interventions,* England: John Wiley & Sons Ltd, 2008; 187-242

22 Jadad AR, Moore RA, Carroll D, *et al*. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials* 1996; 17: 1-12

23 Akaike H. Information Theory and an Extension of the Maximum Likelihood Principle. In: Kotz S and Johnson NL, eds. *in Breakthroughs in Statistics volume 1.* Springer Verlag, US 1992; 610-624

24 Abdelhamid AS, Brown TJ, Brainard JS, *et al*. Omega-3 fatty acids for the primary and secondary prevention of cardiovascular disease. *Cochrane Database Syst Rev* 2020; 3: CD003177

25 Hayamizu K, Aoki Y, Izumo N, Nakano M. Estimation of inter-individual variability of protein requirement by indicator amino acid oxidation method, *J Clin Biochem Nutr* 2021; 68, 32-36

26 DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986; 7: 177-188

27 Hothorn T and Everitt BS. *A Handbook of Statistical Analyses using R THIRD EDITION.* FL: CRC press, 2014; 313-330

Chapter 3.

Safety assessment of L-Arg oral intake in healthy subjects: a systematic review of randomized control trials

3.1 BACKGROUND

L-Arginine (L-Arg) is a nonessential amino acid, but it plays many roles in the body. It is thought to be especially necessary for long-term growth and during development[1]. Accordingly, L-Arg is also called a "conditionally essential amino acid." It is involved in protein synthesis as a proteinogenic amino acid and in the detoxification of ammonia as a constituent of the urea cycle. It also produces nitric oxide (NO) as a precursor of L-citrulline (L-Cit)[2,3]. Moreover, it has been reported that NO promotes blood flow[4] and improves male sexual function[5], indicating NO's vasodilatory activity. NO has also been reported to elevate growth hormone activity[3,6,7] and enhance wound healing[3,7,8]. Consequently, NO is occasionally used for improving athletic performance[2]. For these reasons, L-Arg is often used as a dietary supplement, especially in sports nutrition, all over the world. L-Arg is also known for stimulating insulin secretion and is used as a non-glucose secretagogue (a drug that causes or stimulates secretion) to measure insulin secretion levels[9]. L-Arg is now being used in various fields for human health, but there is only limited information about its safety upon overdose.

From the results of a nonclinical study using rats, Blachier et al. reported that the NOAEL of L-Arg is 3318 mg/kg BW/day in males and 3879 mg/kg BW/day in females[10]. The acceptable daily intake (ADI) for humans can be calculated using these data based on a safety factor (SF) of 100 as follows:

$$ADI = SF \times NOAEL = 1/10 \times 1/10 \times 3318\ (mg/kg) = 3.318(mg/kg)$$

This calculation is based on the body weight of adult men of 60 kg is 199.08 (mg/body/day). However, the mean intake of L-Arg from meals in adults is reported to be 59.71 mg/kg BW/day [10]. Thus, when the body weight is 60 kg, the corresponding intake is 3582.6(mg/body/day). Trumbo et al. reported that the mean intake of dietary L-Arg is 4.2 g/day, while its 99th percentile is 10.1 g/day[11]. In addition, Iguacel et al. reported that the mean intake is 3.6 g/day, while the

interquartile range is 2.8–4.3 g/day[12]. The estimated dietary intake level thus greatly surpasses the ADI and it is not realistic.

The NOAEL of L-Arg in healthy young people with a body weight is 70 kg has been estimated to be 20–40 g/day[13]. This value is estimated from the conversion table of the FDA by using the dosage for which no toxic findings were obtained in a nonclinical trial in which rats or pigs were administered L-Arg over 90 days[14,15]. Blachier et al. evaluated the safety of L-Arg using the NOAEL/UC ratio, based on the two parameters of NOAEL and usual consumption (UC) [10]. This ratio in humans and rats is 4.9 and 9.1, respectively, indicating that the range of safe use in humans is narrow compared with that in animals. The dose calculated from the conversion table is used as evidence for deciding on the initial amount such as medicine or food in a phase I study where administered them to humans for the first time. Therefore, it isn't enough to decide this safe upper intake of L-Arg by only using the conversion table.

McNeal et al. conducted a double-blind placebo-controlled trial targeting overweight but otherwise healthy subjects aimed at clarifying the tolerability of L-Arg and reported that there were no problems associated with L-Arg at 30 g/day for 3 months[13,16]. However, this test involved a wash-in period in which L-Arg was administered at 12 g/day for 1 week before group allocation and excluded subjects who dropped out during this period. We found that adverse events (AEs) had already occurred at 12 g/day because three subjects with intolerable gastrointestinal symptoms dropped out during the wash-in period. For this reason, the report actually focused on the results for subjects highly tolerant of L-Arg, so the generally safe amount of L-Arg remained unclear.
Blachier et al. and Elango published reviews on the NOAEL of L-Arg in humans[10,17]. However, they also calculated the value by quoting the result of McNeal et al., so their NOAEL is probably lower than the reported value.

In recent years, systematic review (SR) has been used as an approach for assessing safety as well as effectiveness and usefulness of amino acids. However, to the best of our knowledge, no SR assessing the safety of L-Arg has been reported yet. For this reason, we designed a study to assess the safety of L-Arg by using SR.
Shao reported that the OSL for L-Arg is 20 g/day[2]. However, we think that an additional approach is needed to confirm this reported dosage as the reliable safe upper limit.

The hockey stick model is one of the response models used in pursuit of the threshold for the expression of toxicity in toxicological studies[18]. This model is characterized by a biphasic pattern with one area reflecting no dose response and the other area reflecting a dose response. We

possibly find the true threshold by using this model due to the threshold dose (TD) which toxicity is expressed is considered to be in between NOAEL and LOAEL.

We proposed a change-point regression model (CPRM) as an analytical method that includes the above-mentioned threshold in it. We also reported the usefulness of weighted CPRM (w-CPRM) weighted by the heterogeneity between clinical studies used in an SR added to this model[19]. CPRM has been applied to estimate the required amounts of essential amino acids and proteins[20,21], but to the best of our knowledge no examples of its application to assessing the safety of amino acids in SR have been reported.

In this study, we performed an SR including only randomized double-blind controlled trials as intervention trials, given that they feature a control group, and conducted an assessment of the safety of L-Arg intake in healthy people using the occurrence of gastrointestinal symptoms as an index. Specifically, we applied w-CPRM and considered the TD based on the results of 31 studies (including 35 tests) as targets of SR.

3.2 MATERIAL AND METHODS

We conducted an assessment of the safety of L-Arg intake targeting healthy people by SR. We set the eligibility criteria of the study (PICOS) as follows: patients (P), healthy people who took L-Arg orally; intervention (I), L-Arg; comparison (C), placebo; outcome (O), any adverse events; and study design (S), intervention trial (randomized double-blind controlled trial).

We followed the Cochrane Handbook for Systematic Reviews of Interventions in conducting this SR and meta-analysis[22]. The results are reported in accordance with the PRISMA 2020 statement: updated guidelines for reporting systematic reviews[23]. The review protocol was registered at umin.ac.jp as UMIN000046133 before the beginning of the study.

3.2.1 Study selection

A systematic search was performed using the PubMed, Cochrane Library, EBSCOhost, and Ichushi-Web databases to identify reports on studies involving L-Arg intervention in humans published until May 2021. Search terms included “L-Arg,” “double-blind,” and “randomized controlled trial.” To investigate all adverse events observed during the intervention trial, we included all oral L-Arg intervention studies without placing restrictions regarding background factors, environment, and sample size. Manual searches of journal articles and reference lists from relevant publications were performed to ensure that all appropriate studies were considered for inclusion. Unoriginal studies and duplicates were removed. Two investigators performed the electronic search independently. Papers were chosen by title and contents of the abstract in the first screening, after which the full text was read. Then, papers describing studies in which L-Arg intervention was performed were selected in the second screening and finally those that matched the PICOS criteria were adopted.

3.2.2 Inclusion and exclusion criteria

Studies identified from the systematic search were included or excluded according to the following criteria. The inclusion criteria were as follows: 1) study on healthy humans, 2) L-Arg administered orally, 3) L-Arg or L-Arg HCl forms as the intervention samples, and 4) double-blind randomized controlled trial. The exclusion criteria were as follows: 1) study design other than an L-Arg intervention study, 2) non-oral administration route of L-Arg, 3) L-Arg was not administered alone, 4) non-healthy humans, 5) L-Arg used as a salt for other acidic drugs, 6) meta-analysis, and 7) no information about the L-Arg dose.

3.2.3 Data extraction

Two investigators independently extracted the following data from eligible papers: 1) name of

the first author, 2) year of publication, 3) study location, 4) study design, 5) numbers of participants in the L-Arg and control groups, 6) participant age, 7) L-Arg dosage per one-time and day, 8) duration of administration of L-Arg, and 9) AEs during the period of L-Arg treatment. Regarding the dosage, when L-Arg was used for intervention as a hydrochloride, it was converted to net L-Arg content. When information was ambiguous or missing, we contacted the corresponding author to obtain the most accurate data available.

3.2.4 Methodological quality

Assessment of the quality of studies was conducted using the Cochrane Collaboration tool for assessing the risk of bias (RoB)[22] and Jadad score[24]. The Cochrane Collaboration tool for assessing the risk of bias includes random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data, selective reporting, and other sources of bias. Studies were classified as having high, low, or unclear risk of bias, according to each criterion.

Jadad score reflects five issues: was the study described as randomized, the randomization was adequate or not, was the study described as double-blind, the double blinding was adequate or not, and was there a description of withdrawals and drop outs. Jadad score was obtained by adding these results based on answers of "yes" = +1 and "no" = 0 (max. score 5). A score of ≥3 was considered to reflect relatively high quality.

3.2.5 Classification of included papers

The included papers were categorized into the following three categories: Category A, with no descriptions of AEs in the article; Category B, stating that no AEs occurred; and Category C, mentioning that AEs occurred during the trial[25].

3.2.6 Estimation of the TD[19]

The exploration of the threshold of L-Arg dosage for estimating the TD was conducted with w-CPRM to take into account the heterogeneity of each study, which is one of the problems specific to SR. w-CPRM is defined as follows:

$$\omega_i y_i = \omega_i \beta \mathrm{I}(\chi_i > \chi_{cp})(\chi_i - \chi_{cp}) + \omega_i \varepsilon_i$$

where y_i is the risk difference (RD) and x_i is the dosage of L-Arg. $I(\cdot) = 1$ if $x_i > x_{cp}$ and is 0 otherwise. In this study, the weight of each study's RD in the random effects model is ω_{1i}, which is included in the formula for w-CPRM as shown below:

$$\omega_{1i} y_i = \omega_{1i} \beta \mathrm{I}(\chi_i > \chi_{cp})(\chi_i - \chi_{cp}) + \omega_{1i} \varepsilon_i \qquad (Eq.1)$$

ω_{1i} is calculated as follows:

$$\omega_{1i} = \frac{1}{SE_i^2 + \widehat{\tau^2}}$$

Here, SE_i is the standard error of RD of each study, while τ^2 denotes variability among the true effects that sample characteristics may introduce[26].

We considered the TD by using w-CPRM including the above-mentioned weight (ω_{1i}). Comparison of each model and setting of the dosage threshold were conducted by the maximum likelihood method used Akaike's information criterion (AIC) as an indicator[27].

Safety was assessed using the frequency of AEs in the L-Arg and control groups. The variable modified RD and 95% CIs were further used to calculate pooled risk estimates. Cochran's Q tests and I^2 statistics were used to examine heterogeneity between studies. The random effects model was used for data synthesis. Sensitivity analysis was performed to identify any study responsible for heterogeneity and/or to test the validity of the conclusions by omitting one study sequentially (leave-one-out approach). Publication bias or small study effect was assessed by the funnel plot method and using Egger's test[28]. The meta-analysis and summary of the risk of bias were conducted using Cochrane Program Review Manager (RevMan) version 5.4[29]. Statistical analysis was conducted using R4.1.1[30] with the packages "**meta**" and "**metafor**" [26,31].

3.2.7 Weighting for meta-analysis considering RoB assessment results

The result of RoB assessment is considered to one of heterogeneities, and if it is, we need to add them to weights (see Chapter 2). ω_{2i} is as follows.

$$\omega_{2i} = \omega_{1i} + RS_i = \frac{1}{SE_i^2 + \widehat{\tau^2}} + RS_i$$

$$RS_i = \frac{score_i}{(\sum_{i=1}^{n} score_i)}$$

RS_i is derived using the results of the RoB assessment for each trial. The RoB evaluation by the Cochrane Collaboration is rated on three levels, low, high, and unknown, for each bias risk item. We scored these results as "low risk" = 1, "high risk" = –1, and unclear = 0 and set $score_i$ as the value upon summing them up.

Adjusted meta-analysis method with RoB has been reported although the number of such studies is limited. Turner et al. have proposed a method for scoring the risk of bias and using it as a correction value for the weighting in meta-analyses[32]. However, due to its specific and complicated, this method has not been widely adopted and is not used for risk assessment by the Cochrane Collaboration[33]. Additionally, Morizane has proposed a method to adjust bias in meta-analysis using Chalmers's score, an evaluation approach for Randomized Control Trials (RCTs) that is not commonly employed[34,35].

Chalmers et al. proposed a scoring system out of 100 to evaluate RCTs. They adjust bias by multiplying the value obtained by dividing the sum of scores in each item by the full mark (100 in this case) to the weighting.

Therefore, using the results of RoB of The Cochrane Collaboration, which is used as standard, we examined a method to calculate the weighting as ω_{3i} as a modification of the method of Morizane as follows.

$$\omega_{3i} = \omega_{1i} * RS_i = \frac{1}{SE_i^2 + \widehat{\tau^2}} * RS_i$$

$$RS_i = \frac{score_i}{14}$$

The Cochrane assessment has seven items in total, so we regarded a perfect score as 14, with high quality of items as +2, unclear as +1, and low quality as 0. We regarded the total score of each study as $score_i$ and we regarded RS_i as $score_i$ divided by 14 (reflecting full marks), which we used to adjust ω_1.

3.2.8 Certainty of evidence

We assessed the certainty of evidence according to the Grading of Recommendations Assessment, Development and Evaluation (GRADE). This was based on four grades, namely, high quality, moderate quality, low quality, and very low quality, for the following five items: study limitations, inconsistency of results, indirectness of evidence, imprecision, and reporting bias[36].

3.3 RESULTS

Of the 445 papers retrieved through electronic and manual searches, 307 papers were excluded after screening of the title and abstract; then, 108 papers were excluded after full-text review. Thus, 34 studies (30 papers) met the inclusion criteria of the review protocol (Fig. 1). These 34 target studies were classified into Categories A–C. Category A included articles not describing the presence or absence of AEs (10 studies). Category B included articles reporting that no occurrence of an AE was observed (17 studies). Category C included articles reporting that the occurrence of an AE was observed (7 studies). The study by McNeal et al.[16] was excluded because they intervened only subjects who were tolerable against for L-Arg and we regarded handling of the AE data was not appropriate.

3.3.1 Characteristics of included studies

The characteristics of the included studies classified into Categories B and C are summarized in Table 1. For the research containing the results of two or more studies reported in the same paper, we listed the results separately. The proportion of men ranged from 0 to 100%, with a mean age of the subjects in each study was 10–73.8 years. The dosage of L-Arg tested in the studies ranged from 2000 to 30,000 mg/day (or / one-time dose), and the duration of administration ranged from 1 to 84 days.

3.3.2 Description of study quality

The included studies varied in terms of their quality (Fig. 2). In all target study, only one in each item except of "other bias" was evaluated as high risk of bias and others were evaluated low or unclear risk of bias. Nine studies were evaluated high risk of bias to "other bias". In most cases, the high risk of bias was due to support such as supplies and funds being provided by a company. Other reasons for high risk of bias included the exclusion of results from timepoints in which it was considered that L-Arg could not be effective and there being an insufficient wash-out period in a cross-over trial. In terms of the Jadad score, 3 studies had a score of ≤2, while 31 studies had a score of ≥3. In addition, among the 24 studies corresponding to Categories B and C, all studies were of high quality with a score of ≥3, with one exception in which the study had a score of 2 (Table 1).

3.3.3 Maximum dose, duration of administration, and sample size

Figure 3 provides a summary of the studies included in this SR as bubble plots. The x- and y-axes indicate the duration of administration and the dosage of L-Arg, respectively. The size of the bubble indicates the sample size of the study. Figure 3-a shows a breakdown of the 34 studies targeted for evaluation. Meanwhile, Figure. 3-b shows the 24 studies classified into Categories B

and C in which the occurrence or absence of AEs could be judged. We set the primary endpoint as gastrointestinal symptoms because these have been reported most among the AEs.

Gastrointestinal symptoms were defined as nausea, vomiting, diarrhea, and abdominal problems such as cramps and bloating. We decided that it is crucial to consider the use of one-time dose because gastrointestinal symptoms are considered a relatively acute response. Therefore, we showed L-Arg dose per day and per one-time, respectively. The x-axis in Fig. 3-a and -b indicates the daily dose of L-Arg, while that in Fig. 3-c indicates the one-time dose. Figure. 3-c shows the breakdown of 23 studies with the exception of that by Vignini et al.[37] because it did not report the dosage per one-time administration. Among the studies in Category A, the highest dose was 14,200 mg/person/day (there was no description of how many doses L-Arg was divided into) and the longest duration of administration was 180 days (both the highest and the longest doses were administered by Fricke et al.[38]). Meanwhile, the largest sample size was 30[38,39]. Reports on studies within Category A did not describe the presence or absence of AEs; therefore, the information obtained from Category A was the longest dosing period and dose of L-Arg in humans as an amount added to an ordinary diet.

In studies within Category B, the highest dose and longest duration of administration were 30,000 mg/day[40] and 84 days[41], respectively. The highest one-time dose was 10,000 mg (20,000 mg as a daily dose)[42].

In Category C, the highest dose and the longest duration of administration were 30,000 mg/day [43] and 45 days[44], respectively. The highest one-time dose was 30,000 mg[43].

Since the reports on the studies within Categories B and C described the presence and absence of AEs, we could use them to estimate the safety of L-Arg. From the Category B and C data, the maximum dose of L-Arg with no AEs was 30,000 mg/person/day as an amount added to an ordinary diet. Administration was performed for 8 days and the L-Arg dosage was divided up into four administrations (7500 mg/person/ one-time dose)[40]. Meanwhile, the minimum dose at which an AE apparently caused by L-Arg was observed was 2000 mg/person/day in a single dose[44].

However, the RD of the rate of occurrence of gastrointestinal symptoms was 0.04 (95% CI: −0.06 to 0.13), which was not statistically significant (P=0.45). Therefore, 6105 mg/person/day in a single dose, which was the second highest reported dose within Category C, became LOAEL[45]. For that reason, NOAEL was estimated at 6000 mg/person/ one-time dose. We estimated the TD of a daily dose as 12,000 mg/day (4000 mg t.i.d), which is under the LOAEL and is the highest dose in Category B[46].

3.3.4 Meta-analysis of AEs

We conducted a meta-analysis on the risk of AEs associated with L-Arg administration using the studies classified within Categories B and C (Fig. 4). The study by Vignini et al.[37] was excluded because it did not report the amount of one-time L-Arg dose; thus, 23 studies were a target for the

meta-analysis within Categories B and C. RD regarding the occurrence of gastrointestinal symptoms reported by Savoye et al.[43] was 0.63 (95% CI: 0.27–0.98), which was significant (P=0.0005). However, upon integrating all of the targeted studies, RD was 0.01 (95% CI: −0.02 to 0.04), indicating no significant increase in gastrointestinal symptoms in association with L-Arg (P=0.34). The heterogeneity among the studies was also small (I^2=9%). Besides, we separated one-time dose into four levels (<3000 mg, ≤3000 to <6000 mg, ≤6000 to >9000 mg, and ≥9000 mg), and considered the dose response by stratified analysis. RDs were 0.02 (95% CI: −0.05 to 0.08), 0.00 (95% CI: −0.03 to 0.03), 0.01 (95% CI: −0.05 to 0.08), and 0.16 (95% CI: −0.04 to 0.35), respectively, revealing a weak dose response. However, there were no significant differences in each subgroup. From the leave-one-out approach that was conducted for sensitivity analysis, slight changes in both χ^2 value and I^2 value were observed, but there were not seriously impact comparing to the overall effect. Therefore, there were no studies that were sufficiently heterogeneous to influence the integration result (see APPENDIX Table S1). In addition, no evidence of publication bias was identified from the funnel plot (Fig. 5) or Egger's test (P=0.1676).

3.3.5 Estimation of the TD by w-CPRM

There are sometimes no threshold exists on the expression of AE depending on a kind of chemical. For that reason, we verified whether the threshold exists on the expression of AE to find the TD of L-Arg at first. The feasibility of applying CPRM was confirmed by comparing it to a linear regression model with AIC as an index. Here, Models A and B were established as linear regression models with no weighting and weighting by ω_{1i}, along with Models C and D as CPRM (w-CPRM) with no weighting and weighting by ω_{1i}, respectively.

Eqs. 2–4 are shown below:

$$y_i = \beta\chi_i + \varepsilon_i \qquad (Eq.2, \text{Model A})$$

$$\omega_{1i}y_i = \omega_{1i}\beta\chi_i + \omega_{1i}\varepsilon_i \qquad (Eq.3, \text{Model B})$$

$$y_i = \beta \mathrm{I}(\chi_i > \chi_{cp})(\chi_i - \chi_{cp}) + \varepsilon_i \qquad (Eq.4, \text{Model C})$$

$$\omega_{1i}y_i = \omega_{1i}\beta \mathrm{I}(\chi_i > \chi_{cp})(\chi_i - \chi_{cp}) + \omega_{1i}\varepsilon_i \qquad (Eq.1, \text{Model D})$$

In Fig. 6, the x-axis represents L-Arg dosa per one-time administration, while the y-axis is the RD between the L-Arg group and the placebo group regarding the occurrence of gastrointestinal symptoms. Here, the size of the bubble reflects the weighting. (The sizes of the bubbles of Model A and Model C are certain because there is no weighting.)

In all of Models A to D, these results were statistically significant and the occurrence of AEs was recognized to exhibit a dose-response relationship (Table 2). Next, we recognized that there was a threshold in the dose response regarding the occurrence of AEs because Models C and D fitted better than Models A and B, as shown by comparing their AICs. Model D representing w-CPRM

was the best fitting among the four models. Therefore, the TD by this model was calculated to be 7531 mg/ one-time dose.

3.3.6 Estimation of the TD incorporating RoB assessment results

We compared two weightings, ω_{2i} and ω_{3i} adjusted using the results of the assessment of RoB.

We confirmed whether there was a difference in estimating of the TD depends on how the weighting was adjusted. The AICs using ω_{2i} and ω_{3i} for w-CPRM were -52.10 and -61.77, respectively. The estimated TDs were 10912 mg/one-time dose and 8754 mg/one-time dose, respectively (Table. 3). The threshold dose was lower for ω_{3i} compared to ω_{2i}, but the AIC showed significant improvement.

Additionally, it was confirmed that the determination of TD was obtained from the AIC profile exploring thresholds could be more easily done through a clearer graph (data not shown).

Incidentally, we also checked the results of the linear regression model adjusted by ω_{2i} and ω_{3i}, respectively, and found that L-Arg has TD not only with ω_{1i} but also with ω_{2i} and ω_{3i}.

3.3.7 Certainty of evidence

We confirmed variation in the certainty of evidence from the assessment of the RoB of the 23 studies targeted for the meta-analysis. Random sequence generation and allocation concealment asked to report appropriately due to focus on the existence of AEs in this research. In addition, "other bias" out of the result of RoB stands out for high risk of bias compares with other items but the most of reasons were funding from companies related to materials of L-Arg. The gastrointestinal symptoms that we focus on in this work are considered relatively temporary or tolerable. Therefore, in some reports it is described that the intervention continued until the end of the study even if AEs occurred. As such, it is easy to believe that the gastrointestinal AEs were minor events. In addition, we thought that there were few reports related to AEs among studies targeted here because the researchers had emphasized assessing the effectiveness rather than the safety in view of the fact that L-Arg is one of food components considered safe. For that reason, we thought that there is a serious risk of bias in the outcome that only used studies gathered this time and we judged that it is appropriate to get one level of grade down according to GRADE.

From the result of the meta-analysis about the frequency of gastrointestinal symptoms, I^2 was small as 11% (P=0.31) and we also confirmed the CIs between each study overlap. In the study by Savoye et al.[43], five subjects had gastrointestinal symptoms among eight were reported and we judged that there was a significant difference in the result. Generally, I^2 can be considered to reflect low heterogeneity among studies when it is 40% or less. However, from the results of the leave-one-out analysis, no individual study has an influence that changes the whole outcome of this research. Therefore, we judged that there is no problem with “inconsistency of results”

according to GRADE.

In terms of the inclusion criteria applied in this study, the research by Blum et al.[47,48]. targeted postmenopausal women but healthy. Meanwhile, research by Vignini et al.[37] included subjects with anorexia nervosa (AN) besides healthy subjects, but we selected it here as a target study for inclusion in this research. However, we did not use the data from Vignini et al.[37] in the meta-analysis and w-CPRM because there was no report the amounts of the one-time dose applied. In addition, Savoye et al.[43] administered L-Arg in the stomach through a gastric tube, but this is considered equivalent to oral intake. In this work, we also included studies in which L-Arg intervention involved L-Arg as a hydrochloride, not only L-Arg alone. Nonetheless, we thought there were no differences between them because the main substances themselves were both L-Arg. In terms of the establishment of a control group, outcome, and study design, all studies met the inclusion criteria and there were no points worthy of special mention. Therefore, all studies were considered to adhere to the PICOS guidelines that we set and we judged that there were no problems with “indirectness of evidence” according to GRADE.

Generally, when conducting meta-analyses, a total number of events of 300 or more is preferable in the case where the outcome is a binary variable. However, in this meta-analysis regarding the risk of gastrointestinal symptoms being associated with L-Arg administration, the number of events is below 300 and the 95% confidence interval of RD is included 0. Therefore, it was very serious about "imprecision" and was equivalent to two levels of grade down according to the criteria of GRADE. However, GRADE was devised the assessment of effectiveness of guidelines and so on. Our research is focused on the safety, for that reason, we thought some issues remine to use it directly. Here, TD was calculated as 7531 mg/ one-time dose by w-CPRM. Therefore, we conducted meta-analysis with all studies that used an L-Arg dose of 7531 mg/ one-time dose or more, and we observed a trend of an increasing incidence of AEs (P=0.06). When the outcome of a study is AEs, caution is needed even if a tendency for a significant difference is identified. In a study like the current one, we might not obtain a sufficient total number of events even if we achieve a large sample size because the rate of AEs due to food components is considered low. Therefore, in this research, we considered that it was not particularly essential to fulfill the criteria "reporting bias" according to GRADE.

In terms of the reporting bias, the results of the funnel plot and Egger’s test confirmed that there was no serious influence of such bias. From the assessment of the five items: study limitations, inconsistency of results, indirectness of evidence, imprecision and reporting bias, we eventually judged the certainty of the evidence as being intermediate by considering that there weren’t enough reasons that don't judge no problems about "study limitations" and "imprecision”.

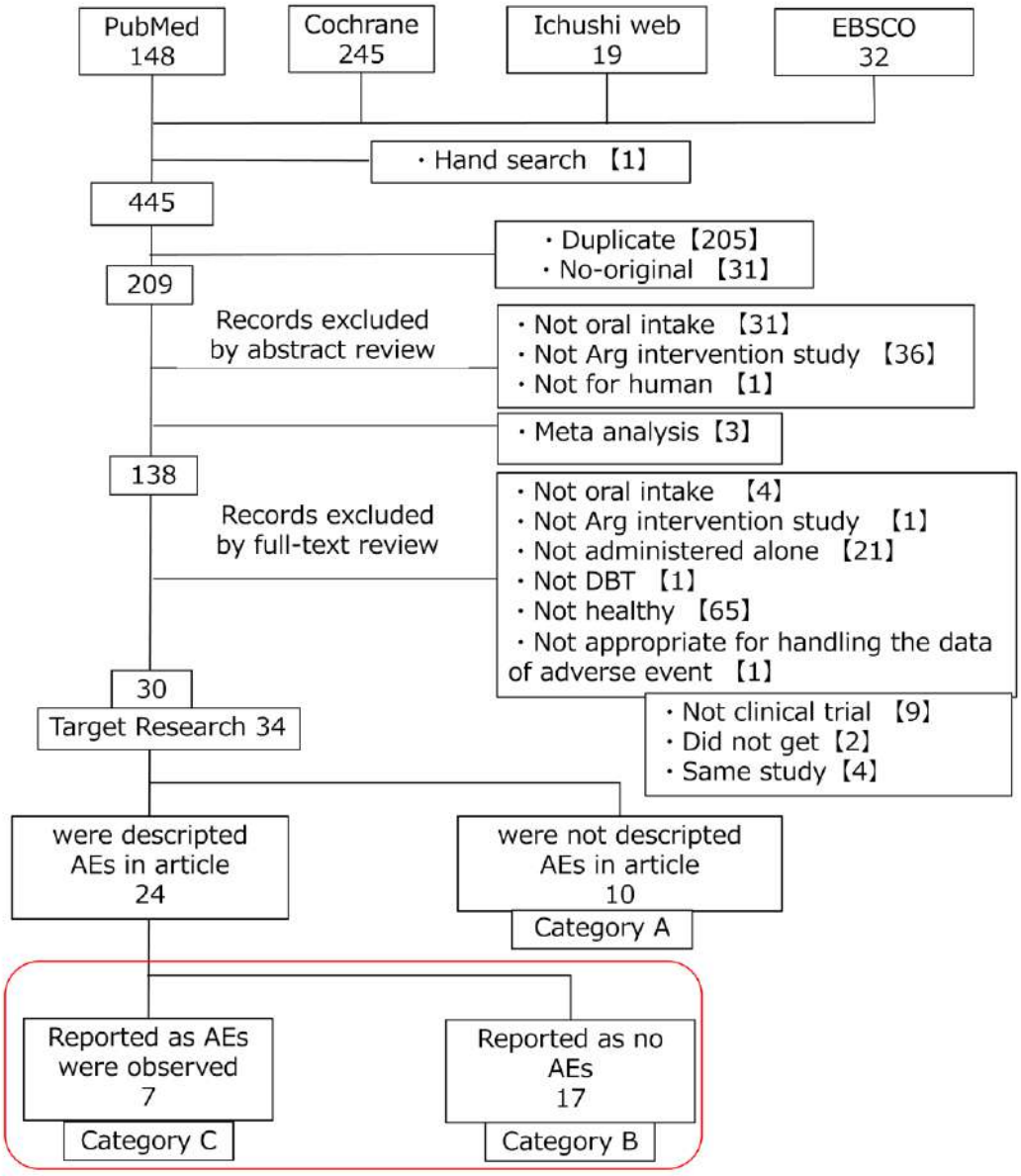

Fig. 1 Flow chart of the selection of studies for inclusion in the systematic review. Articles in Category A did not describe AEs. Articles in Category B described AEs but reported no observed AEs. Articles in Category C reported the occurrence of AEs. AE: adverse event, DBT: double-blind trial.

Table 1 Characteristics of the included studies in systematic review

Study	Country	n_{Arg} / n_{total}	Age (year)	Inclusion criteria	Daily dose (mg)	One-time dose (mg)	Duration of trial (day)	Jadad score	Category
Adams et al. 1995[49]	Australia	12/12	34±1	healthy	21000	7000	3	4	C
Aguiar et al. 2016[50]	Brazil	10/20	71.6±5.9	healthy	8000	8000	1	4	B
Alvares et al. 2012[51,52]	Brazil	ND/15	arg:26.3±4.9, pla:24.7±1.8 (mean±SD)	healthy	4961.5	4961.5	1	4	B
Alvares et al. 2012[51,52]	Brazil	9/17	26±4.6 (mean±SD)	healthy	4961.5	4961.5	1	3	A
Alvares et al. 2014[53]	Brazil	8/15	arg:36.8±7.1, pla:30.6±9.5	healthy	4947.8	1236.95	28	5	A
Andrade et al. 2018[54]	Brazil	10/20	22.8±3.4	healthy	6000	6000	3	3	B
Ast et al. 2011[46]	Poland	7/19	37.9±8.03 (mean±SD)	healthy	6000	2000	28	3	B

ND no data, AN anorexia nervosa

Table 1 Characteristics of the included studies in systematic review (continued)

Study	Country	n_{Arg} / n_{total}	Age (year)	Inclusion criteria	Daily dose (mg)	One-time dose (mg)	Duration of trial (day)	Jadad score	Category
Ast et al. 2011[46]	Poland	6/19	37.9±8.03 (mean±SD)	healthy	12000	4000	28	3	B
Birol et al. 2019[55]	Turkey	10/19	18.3±0.48 (mean±SD)	healthy	10800	10800	1	3	A
Blum et al. 2000[47,48]	USA	10/10	55±5	Postmenopausal healthy women	9000	3000	30	3	B
Bode-Böger et al. 1998[56]	Germany	8/8	25.2±0.2	healthy	6000	6000	1	3	A
Bode-Böger et al. 2003[57]	Germany	12/12	73.8	healthy	16000	8000	14	3	C
Chin-Dusting et al. 1996[42]	Australia	8/16	arg:21.9±0.6, Pla:20.9±1.0 (mean±SEM)	healthy	20000	10000	28	4	B
Fahs et al. 2009[58]	USA	18/18	24.2±0.7 (mean ±SE)	healthy	7000	7000	1	4	A

ND no data, AN anorexia nervosa

Table 1 Characteristics of the included studies in systematic review (continued)

Study	Country	n_{Arg} / n_{total}	Age (year)	Inclusion criteria	Daily dose (mg)	One-time dose (mg)	Duration of trial (day)	Jadad score	Category
Forbes and Bell 2011[59]	Canada	14/14	25±5	Healthy	5850	5850	1	3	B
Forbes and Bell 2011[59]	Canada	14/14	25±5	Healthy	11700	11700	1	3	C
Forbes et al. 2013[60]	Canada	15/15	28±5 (mean±SD)	Healthy	5850	5850	1	2	B
Forbes et al. 2014[45]	Canada	14/14	25±4 (means±SD)	Healthy	6150	6150	1	4	C
Fricke et al. 2008[38]	Germany	15/30	arg:54.4±4.1, pla:55.3±4.4	Healthy	14200	ND	180	3	A
Luiking et al. 1998[40]	Netherlands	10/10	24.2±4.1	Healthy	30000	7500	8	4	B
Mansoor et al. 2005[61]	Canada	7/7	37.4	Healthy	9923	3307.67	2	4	A
Meirelles and Matsuura 2018[62]	Brazil	12/12	27±3 (mean±SD)	Healthy	4961.5	4961.5	1	3	B
Pahlavani et al. 2014[44]	Iran	28/56	20.85±4.29	Healthy	2000	2000	45	4	C
Robinson et al. 2003[63]	UK	6/6	25±2	Healthy	10000	10000	1	3	C

ND no data, AN anorexia nervosa

Table 1 Characteristics of the included studies in systematic review (continued)

Study	Country	n_{Arg} / n_{total}	Age (year)	Inclusion criteria	Daily dose (mg)	One-time dose (mg)	Duration of trial (day)	Jadad score	Category
Savoye et al. 2006[43]	France	8/8	36	Healthy	15000	15000	1	4	B
Savoye et al. 2006[43]	France	8/8	36	Healthy	30000	30000	1	4	C
Schwedhelm et al. 2008[64]	Germany	20/20	57(NR)	Healthy	3000	1000	7	2	A
Schwedhelm et al. 2008[64]	Germany	20/20	57(NR)	Healthy	3200	1600	7	2	A
Streeter et al. 2019[39]	USA	30/30	20.4±1.8	Healthy	3000	3000	1	4	A
Suzuki et al. 2017[65]	Japan	10/42	20-49	Healthy	2000	2000	7	4	B
Ueno et al. 2018[41]	Japan	15/30	10-17	Healthy	5000	2500	84	4	B
Vanhatalo et al. 2013[66]	UK	18/18	22±3 (mean±SD)	Healthy	6000	6000	1	4	B
Vignini et al. 2010[37]	Italy	40/80	AN:22±5, healthy:23±3	AN, healthy	8300	ND	14	3	B
Vuletic et al. 2013[67]	Croatia	59/117	21.7±1.8	Healthy	3000	3000	1	4	B

ND no data, AN anorexia nervosa

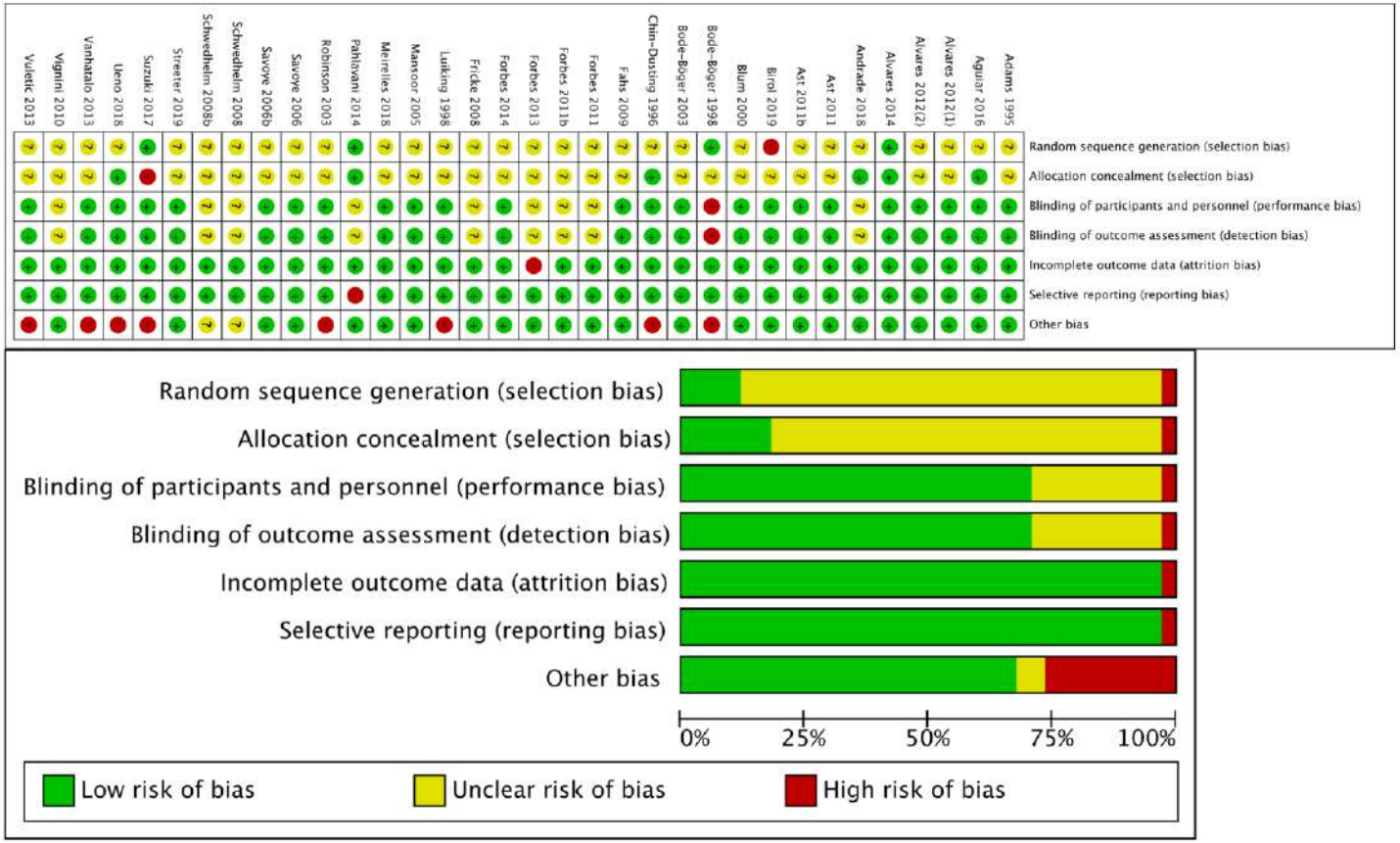

Fig. 2 Assessment of risk of bias for 30 selected human studies: summary of items of bias. Risk of bias for all trials is presented as percentages of trials with low, high, or unclear risk of bias in each assessment item.

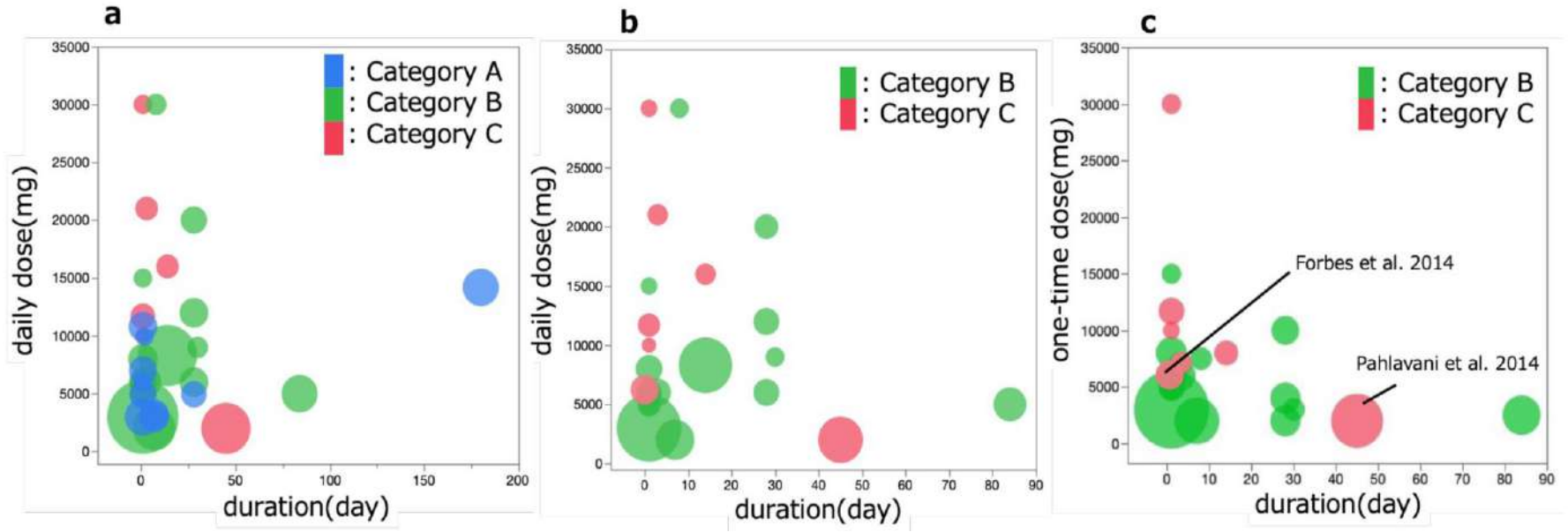

Fig. 3 Summary of all included studies. The x- and y-axes indicate the duration of administration and the dosage of L-Arg, respectively. The size of the bubble indicates the sample size of the study. **Panel a:** all included studies, **panel b:** Categories B and C, which are needed to decide NOAEL and LOAEL, **panel c:** the dose of Categories B and C shown as mg/ one-time dose.

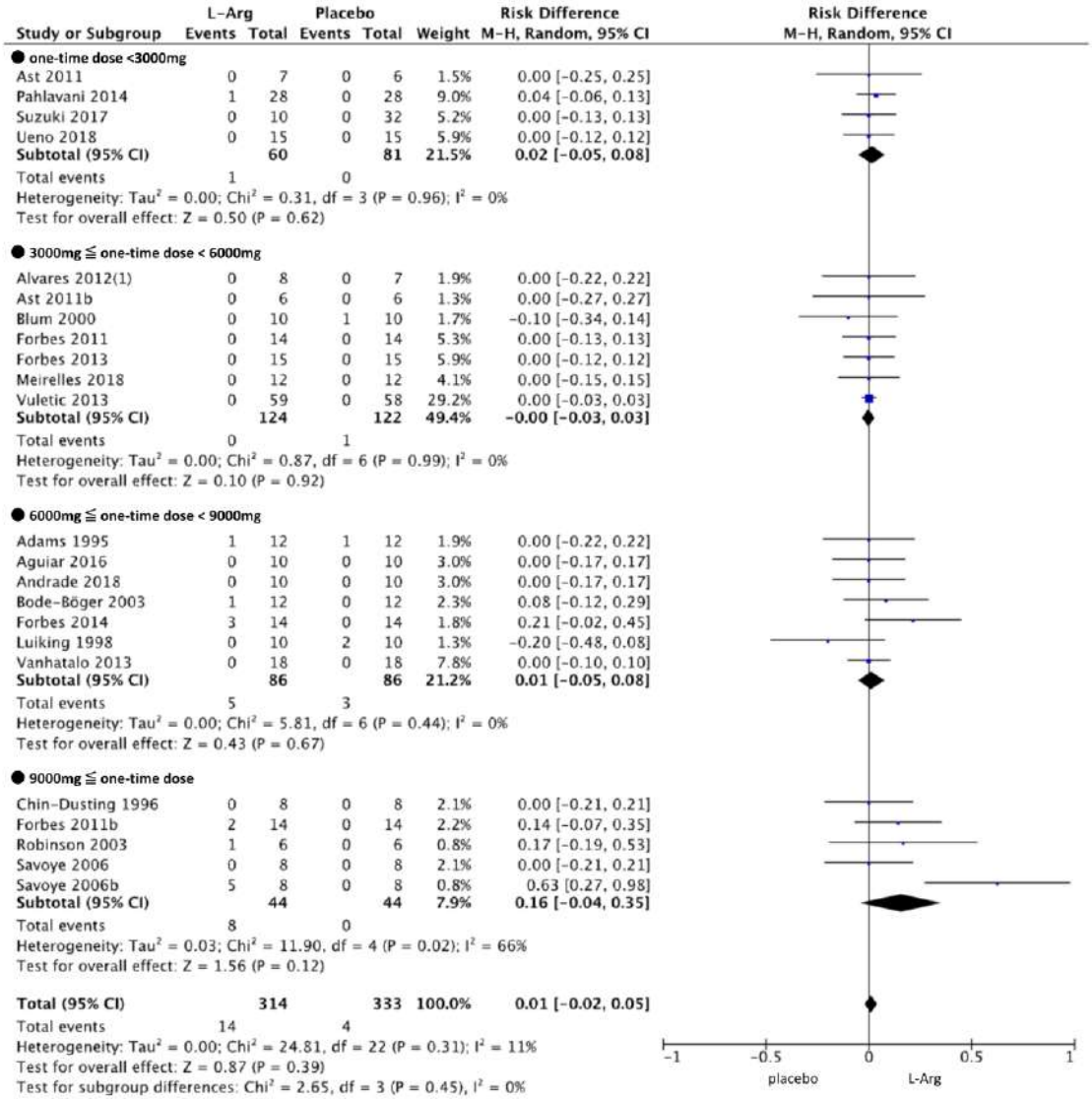

Fig. 4 Difference in risk of suffering gastrointestinal symptoms associated with L-Arg. The included gastrointestinal symptoms were nausea, vomiting, diarrhea, and abdominal problems such as cramp and bloating. Meta-analysis was carried out by stratified analysis based on the four different dose ranges.

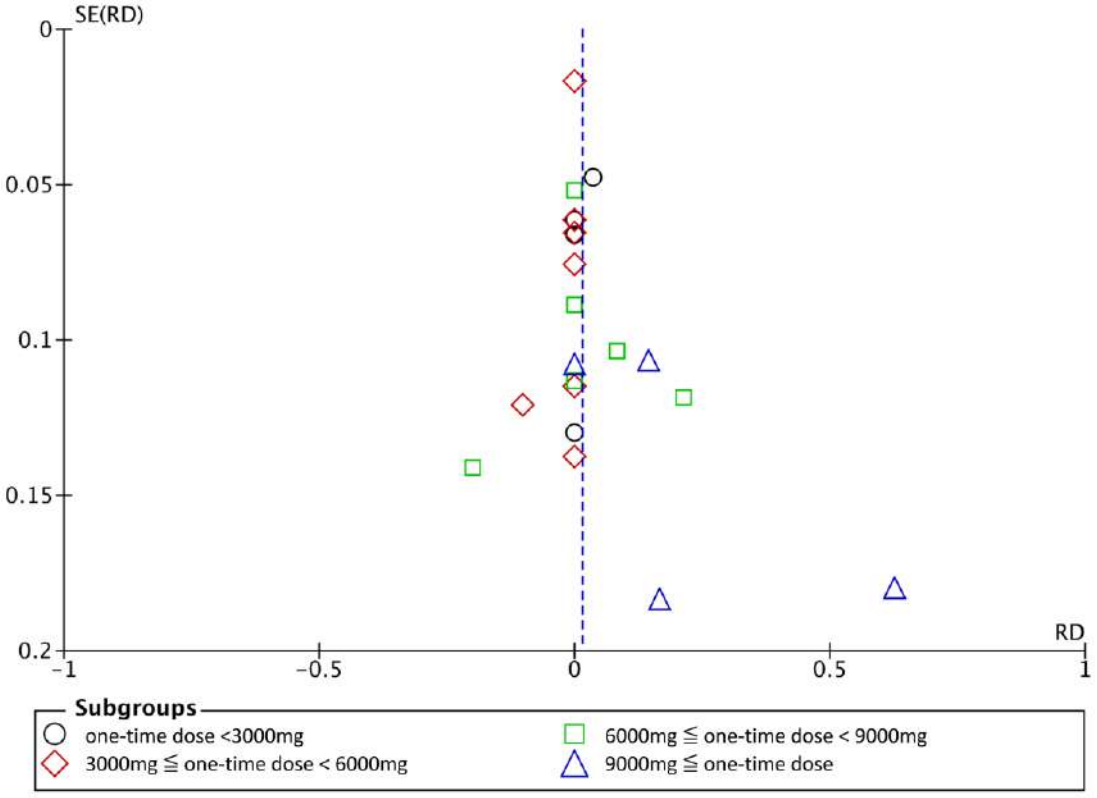

Fig. 5 Results of funnel plot for L-Arg study. There was no evidence of publication bias from the funnel plot or Egger's test (P=0.1676).

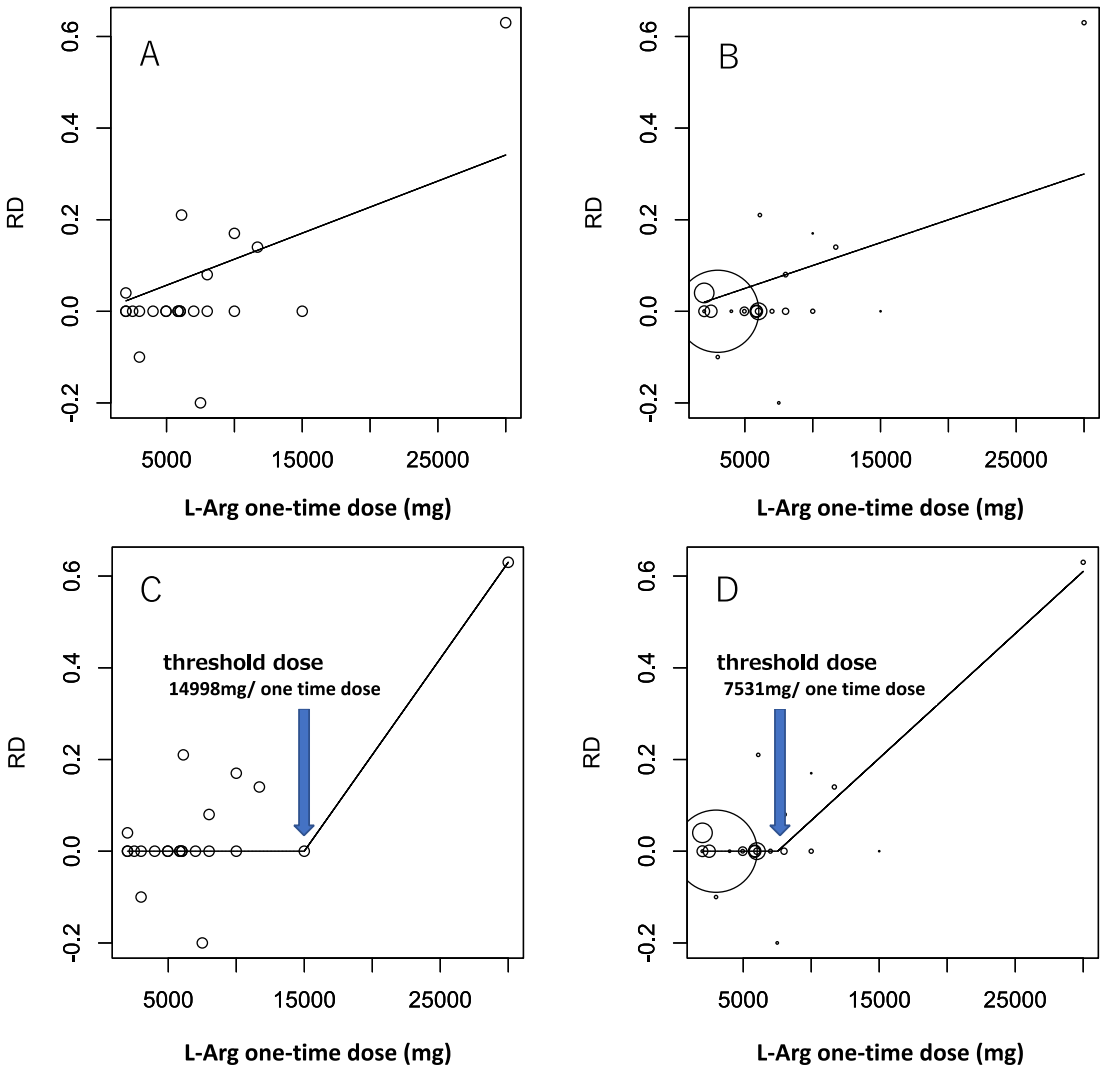

Fig. 6 Results of regression model and change-point regression model (CPRM) for L-Arg study. Panels A and B indicate the results of Eqs. 2 and 3, respectively. Panels C and D indicate the results of Eqs. 1 and 4, respectively. The sizes of the circles in panels B and D indicate weight 1. Panels A and C are drawn by an unweighted model, so the sizes of the circles are the same. RD: risk difference.

Table 2 Summery of models

Model	weight	x_{cp}	AIC_{min}	β	P-value
A ($Eq.$ 2)	-	-	-31.14	1.137e-05	0.000236
B ($Eq.$ 3)	ω_1	-	-38.62	9.991e-06	0.000463
C ($Eq.$ 4)	-	14998	-46.38	4.200e-05	1.33e-07
D ($Eq.$ 1)	ω_1	7531	-62.31	2.716e-05	4.26e-09

Table 3 Summery of w-CPRM with weight 2 and 3

Weight	x_{cp}	AIC_{min}	β	P-value
ω_2	10912	-52.10	3.196e-05	2.06e-08
ω_3	8754	-61.77	2.912e-05	2.5e-09

3.4 DISCUSSION

In this study, we conducted an assessment of the safety of the oral intake of L-Arg based on studies targeting healthy people by using SR. The estimation of the TD was conducted by w-CPRM, which assumed that the incidence of adverse events until reaching the threshold of the expression is 0 and modified the intercept[19].

In our study, gastrointestinal symptoms were the most frequently reported as adverse events, therefore, we conducted by focusing on the occurrence of these events. Previous studies have also reported gastrointestinal symptoms such as nausea and diarrhea as adverse events with L-Arg ingestion[68,69,70]. Adverse events other than gastrointestinal symptoms among the target studies included headache in 2 cases and skin problems in 3 cases (see APPENDIX Table S2). The subject who had a bullous pemphigoid was withdrawn from the study due to the necessity of steroid treatment. However, there was no describe of causal relationships with L-Arg[47,48]. All patients out of the five cases mentioned above dropped out except one case of headache, but there were no other reports of serious adverse events including deaths or hospitalizations. Most of the participants who had gastrointestinal symptoms continued participating until the study ended because these were mild or temporary. No laboratory findings were obtained from the target studies. The study by McNeal was excluded from the target studies of our SR because it was limited only to healthy subjects who tolerated high-dose L-Arg. However, there were no effects on the clinical laboratory data between 0 and 90 days nevertheless the dosage was relatively high: 15000mg/one-time dose (30000mg/day) for 90 days[16].

Among the 34 studies that were targeted for analysis, from the results of meta-analysis on the studies classified into Categories B and C, which reported the existence of AEs, there was no significant association between the reported gastrointestinal symptoms and the intake of L-Arg. There were also no significant results in each group from the subgroup analyses for each singly administered dose category. However, we observed a trend for a dose-dependent increase in RD. Therefore, the rate of occurrence of gastrointestinal AEs might rise upon increasing the one-time dosage of L-Arg, suggesting the need to decide on the upper limit of one-time administered dose.

In the conventional method that was used for the safety assessment of L-Lys, we decided on NOAEL from the studies classified into Categories B and C and then estimated the TD[25]. However, as mentioned in the introduction, this dose is the same as OSL[71]. Therefore, we estimated the TD by using CPRM, which enables to find the exploration of the threshold, as a novel approach this time.

Most meta-analyses are based on sets of studies that are not exactly identical in terms of their

methods and/or the characteristics of the included samples. Such differences may obscure the true effects[26]. Meta-analyses generally uses either the fixed effects model or the random effects model. The fixed effects model is a model that assumes the random errors to the difference between studies and the random effects model is a model that inserts the heterogeneity that can't be ignored the difference between studies to the fixed effects model. Therefore, Meta-analyses uses these errors and differences as weight (ω_{1i}). For that reason, we considered that the same adjustment would also be needed on CPRM. Here, we used the random effects model considering the differences between studies, which is a problem specific to SR. However, τ2 that denotes variability among the true effects that sample characteristics may introduce was 0.00 from the result of Fig. 4 (τ^2=0.03 in the subgroup analysis at doses of 9000 mg or more), and it was interpreted that the heterogeneity among studies was low. Therefore, we thought that the same result could get when using the fixed effects model this time.

In this research, we adopted only RCT, which are generally considered to be the highest-quality study type and assessed the quality of studies. However, variations in both the score of the Cochrane Collaboration tool for assessing the risk of bias[22] and the Jadad score[24] were identified among the studies, suggesting that the studies were not of uniform quality (Tables 1 and Fig. 2). Meanwhile, the RoB results are not normally used as a parameter for meta-analysis. For this reason, we thought to reflect the result of RoB into the w-CPRM as weight to find the TD considering about the heterogeneity of it.

When using CPRM, at least three data points on each of two straight lines and over seven data points as a whole are necessary to accurately estimate the dose. We could judge that sufficient data were obtained to apply CPRM because here the target research for the assessment involved 23 studies. However, when applying CPRM that has no weighting (Model C), the straight line after the dose threshold was limited to two points, with 30,000 mg/ one-time dose being the highest reported dose and 15,000 mg/ one-time dose being the next highest (Fig. 6). Therefore, there were fewer data points on the side of higher dosages, which limited the estimation of the dose. However, the result of the estimation of the TD using this model did not appear to have been markedly impacted because three or more data points were present on both straight lines in w-CPRM weighted by ω_{1i}, ω_{2i} and ω_{3i}.

In this research, there were ten reports corresponding to Category A, which refers to studies with no description of the existence of AEs. We could not use these reports for the meta-analysis and w-CPRM, despite them being target studies for the assessment. As a reason of why these reports in category A exist, foods are generally considered safe. It also seems that one of the reasons why objective assessment of gastrointestinal symptoms is difficult is that there are less severe than

other AEs and cannot be detected by a blood test. Therefore, we thought that there is a limit to estimating the TD using only data from SR.

When the L-Arg dose was 4000 mg/ one-time dose, AIC was −52.83 from the AIC plot of w-CPRM weighted by ω_{1i}. This was the dose estimated using our conventional method to estimate the TD (OSL). AIC was −62.31 when TD was estimated as 7531 mg/ one-time dose which was considered to approach the true value. It was also the same when using w-CPRM weighted by ω_{2i} and ω_{3i}

However, regarding ω_{2i} and ω_{3i}, which include the result of RoB, the estimated dose may vary widely depending on the kinds of RoB and how to insert them into weighting. Therefore, we consider that there is a need for additional research on these issues.

3.5 REFERENCES

1. Laidlaw SA, Kopple JD. Newer concepts of the indispensable amino acids. Am J Clin Nutr. 1987; 46 :593-605
2. Shao A, Hathcock JN. Risk assessment for the amino acids taurine, L-glutamine and L-Arg. Regul Toxicol Pharmacol 2008; 50 :376-99
3. Visek WJ. Arginine needs, physiological state and usual diets. A reevaluation. J Nutr 1986; 116 :36-46
4. Hambrecht R, Hilbrich L, Erbs S, Gielen S, Fiehn E, Schoene N, Schuler G. Correction of endothelial dysfunction in chronic heart failure: additional effects of exercise training and oral L-Arg supplementation. J Am Coll Cardiol 2000; 35 :706-13
5. Barassi A, Corsi Romanelli MM, Pezzilli R, Damele CA, Vaccalluzzo L, Goi G, Papini N, Colpi GM, Massaccesi L, Melzi d'Eril GV. Levels of L-Arg and l-citrulline in patients with erectile dysfunction of different etiology. Andrology 2017; 5: 256-261
6. Chromiak JA, Antonio J. Use of amino acids as growth hormone-releasing agents by athletes. Nutrition 2002; 18: 657-61.
7. Lind DS. Arginine and cancer. J Nutr 2004; 134 :2837S-2841S; discussion 2853S
8. Cheshmeh S, Hojati N, Mohammadi A, Rahmani N, Moradi S, Pasdar Y, Elahi N. The use of oral and enteral tube-fed arginine supplementation in pressure injury care: A systematic review and meta-analysis. Nurs Open 2022; 9: 2552-2561
9. Eremin O. ed. L-Arginine:Biological aspects and clinical application.Chapman & Hall.15-8, 1997
10. Blachier F, Blais A, Elango R, Saito K, Shimomura Y, Kadowaki M, Matsumoto H. tolerable amounts of amino acids for human supplementation: summary and lessons from published peer-reviewed studies. Amino Acids 2021; 53: 1313-1328
11. Trumbo P, Schlicker S, Yates AA, Poos M. Food and Nutrition Board of the Institute of Medicine, The National Academies. Dietary reference intakes for energy, carbohydrate, fiber, fat, fatty acids, cholesterol, protein and amino acids. J Am Diet Assoc 2002; 102 :1621-1630
12. Iguacel I, Perez-Cornago A, Schmidt JA, Van Puyvelde H, Travis R, Casagrande C, Nicolas G, Riboli E, Weiderpass E, Ardanaz E, Barricarte A, Bodén S, Bruno E, Ching-López A, Aune D, Jensen TE, Ericson U, Johansson I, Ma Huerta J, Katzke V, Kühn T, Sacerdote C, Schulze MB, Skeie G, Ramne S, Ward H, Gunter MJ, Huybrechts I. Evaluation of protein and amino acid intake estimates from the EPIC dietary questionnaires and 24-h dietary recalls using different food composition databases. Nutr Metab Cardiovasc Dis. 2022; 32:80-89
13. Cynober L, Bier DM, Kadowaki M, Morris SM Jr, Elango R, Smriga M. Proposals for Upper Limits of Safe Intake for Arginine and Tryptophan in Young Adults and an Upper Limit of

Safe Intake for Leucine in the Elderly. J Nutr 2016; 146: 2652S-2654S
14. Wu Z, Hou Y, Hu S, Bazer FW, Meininger CJ, McNeal CJ, Wu G. Catabolism and safety of supplemental L-Arg in animals. Amino Acids 2016; 48: 1541-1552
15. Guidance for Industry Estimating the Maximum Safe Starting Dose in Initial Clinical Trials for Therapeutics in Adult Healthy Volunteers (2005) https://www.fda.gov/media/72309/download Accessed 9 May 2023
16. McNeal CJ, Meininger CJ, Wilborn CD, Tekwe CD, Wu G. Safety of dietary supplementation with arginine in adult humans. Amino Acids 2018; 50: 1215-1229
17. Elango R. Tolerable Upper Intake Level for Individual Amino Acids in Humans: A Narrative Review of Recent Clinical Studies. Adv Nutr 2023; 14 :885-894.
18. Johnson GE, Doak SH, Griffiths SM, Quick EL, Skibinski DO, Zaïr ZM, Jenkins GJ. Non-linear dose-response of DNA-reactive genotoxins: recommendations for data analysis. *Mutat Res* 2009; 678: 95-100
19. Kuramochi Y and Hayamizu K. A change-point regression approach for estimating no observed adverse effect level from systematic review. J Clin Biochem Nutr 2023; 72 :256-262.
20. Hayamizu K, Kato M, Hattori S. Determining amino acid requirements from repeated observations on indicator amino acid oxidation method by mixed-effect change-point regression models. J Clin Biochem Nutr 2011; 49: 115-20
21. Hayamizu K, Aoki Y, Izumo N, Nakano M. Estimation of inter-individual variability of protein requirement by indicator amino acid oxidation method. J Clin Biochem Nutr 2021; 68: 32-36
22. Higgins JPT, Green S (2008) Cochrane Handbook for Systematic Reviews of Interventions. Chichester: Wiley-Blackwell
23. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, Chou R, Glanville J, Grimshaw JM, Hróbjartsson A, Lalu MM, Li T, Loder EW, Mayo-Wilson E, McDonald S, McGuinness LA, Stewart LA, Thomas J, Tricco AC, Welch VA, Whiting P, Moher D. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021; 372:n71
24. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, McQuay HJ. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Control Clin Trials 1996; 17: 1-12
25. Hayamizu K, Oshima I, Fukuda Z, Kuramochi Y, Nagai Y, Izumo N, Nakano M. Safety assessment of L-lysine oral intake: a systematic review. Amino Acids 2019; 51 :647-659
26. Viechtbauer and Wolfgang. Conducting Meta-Analyses in R with the metafor Package. *Journal of Statistical Software* 2010; 36: 3
27. Akaike H. Information Theory and an Extension of the Maximum Likelihood Principle.

Breakthroughs in Statistics, Vol.I, Foundations and Basic Theory, S. Kotz and N.L.Johnson, eds., Springer-Verlag, 1992; 610-624

28. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ. 1997; 315(7109):629-634.
29. The Cochrane Collaboration (2020) https://revman.cochrane.org/info. Accessed 9 May 2023
30. The R Project (2021) https://www.r-project.org. Accessed 9 May 2023
31. Schwarzer G, Carpenter JR, Rücker G. *Meta-Analysis with R.* Springer International Publishing, 2015; 85-104
32. Turner RM, Spiegelhalter DJ, Smith GC, Thompson SG. Bias modelling in evidence synthesis. J R Stat Soc Ser A Stat Soc. 2009 Jan; 172 (1): 21-47.
33. Toshio M. bias ni yoru koukashihyou no tyousei. Imic. 2016; 37 (1): 8-18
34. Toshio M. 2005 July. Wakariyasui igakutoukeigaku. Tokyo: Medical Tribune, Inc. 50-58
35. Chalmers TC, Smith H Jr, Blackburn B, Silverman B, Schroeder B, Reitman D, Ambroz A. A method for assessing the quality of a randomized control trial. Control Clin Trials. 1981 May; 2(1): 31-49.
36. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, Schünemann HJ. GRADE Working Group. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008; 336(7650):924-926.
37. Vignini A, D'Angelo M, Nanetti L, Camilloni MA, Cester AM, Faloia E, Salvolini E, Mazzanti L. Anorexia nervosa: a role for L-Arg supplementation in cardiovascular risk factors? Int J Eat Disord 2010; 43 :464-71
38. Fricke O, Baecker N, Heer M, Tutlewski B, Schoenau E. The effect of L-Arg administration on muscle force and power in postmenopausal women. Clin Physiol Funct Imaging 2008; 28 :307-311
39. Streeter DM, Trautman KA, Bennett TW, McIntosh LE, Grier JW, Stastny SN, Hackney KJ. Endothelial, Cardiovascular, and Performance Responses to L-Arg Intake and Resistance Exercise. Int J Exerc Sci 2019; 12:701-713
40. Luiking YC, Weusten BL, Portincasa P, Van Der Meer R, Smout AJ, Akkermans LM. Effects of long-term oral L-Arg on esophageal motility and gallbladder dynamics in healthy humans. Am J Physiol 1998; 274 :G984-991
41. Ueno K, Oonuki K, Goto J, Sato D, Shimizu K, Ohnuki K. Safety of Intake of L-Arg Containing Hight Increase Supplements in Healthy Children and Adolescents: A Randomized, Double-blinded, Placebo-controlled Trial. Journal of New Remedies & Clinics 2018; 67:1075-1084
42. Chin-Dusting JP, Alexander CT, Arnold PJ, Hodgson WC, Lux AS, Jennings GL. Effects of in vivo and in vitro L-Arg supplementation on healthy human vessels. J Cardiovasc Pharmacol 1996; 28 :158-166

43. Savoye G, Jemaa Y, Mosni G, Savoye-Collet C, Morcamp P, Déchelotte P, Bouin M, Denis P, Ducrotté P. Effects of intragastric L-Arg administration on proximal stomach tone under basal conditions and after an intragastric diet. Dig Dis Sci 2006; 51:2147-2153
44. Pahlavani N, Jafari M, Sadeghi O, Rezaei M, Rasad H, Rahdar HA, Entezari MH. L-Arg supplementation and risk factors of cardiovascular diseases in healthy men: a double-blind randomized clinical trial 2014; 3:306
45. Forbes SC, Harber V, Bell GJ. Oral L-Arg before resistance exercise blunts growth hormone in strength trained males. Int J Sport Nutr Exerc Metab 2014; 24 :236-244
46. Ast J, Cieślewicz AR, Korzeniowska K, Bogdański P, Kazmierczak E, Olszewski J, Skołuda A, Jabłecka A. Supplementation with L-Arg does not influence arterial blood pressure in healthy people: a randomized, double blind, trial. Eur Rev Med Pharmacol Sci 2011; 15 :1375-1384
47. Blum A, Cannon RO 3rd, Costello R, Schenke WH, Csako G. Endocrine and lipid effects of oral L-Arg treatment in healthy postmenopausal women. J Lab Clin Med 2000a; 135: 231-237
48. Blum A, Hathaway L, Mincemoyer R, Schenke WH, Kirby M, Csako G, Waclawiw MA, Panza JA, Cannon RO 3rd. Effects of oral L-Arg on endothelium-dependent vasodilation and markers of inflammation in healthy postmenopausal women. J Am Coll Cardiol 2000b; 35 :271-276
49. Adams MR, Forsyth CJ, Jessup W, Robinson J, Celermajer DS. Oral L-Arg inhibits platelet aggregation but does not enhance endothelium-dependent dilation in healthy young men. J Am Coll Cardiol 1995; 26 :1054-1061
50. Aguiar AF, Balvedi MC, Buzzachera CF, Altimari LR, Lozovoy MA, Bigliassi M, Januário RS, Pereira RM, Sanches VC, da Silva DK, Muraoka GA. L-Arg supplementation does not enhance blood flow and muscle performance in healthy and physically active older women. Eur J Nutr 2016; 55 :2053-2062
51. Alvares TS, Conte CA, Paschoalin VM, Silva JT, Meirelles Cde M, Bhambhani YN, Gomes PS. Acute L-Arg supplementation increases muscle blood volume but not strength performance. Appl Physiol Nutr Metab 2012a; 37:115-126
52. Alvares TS, Conte-Junior CA, Silva JT, Paschoalin VM. Acute L-Arg supplementation does not increase nitric oxide production in healthy subjects. Nutr Metab (Lond) 2012b; 9:54
53. Alvares TS, Conte-Junior CA, Silva JT, Paschoalin VM. L-Arg does not improve biochemical and hormonal response in trained runners after 4 weeks of supplementation. Nutr Res 2014; 34 :31-39
54. Andrade WB, Jacinto JL, da Silva DK, Roveratti MC, Estoche JM, Oliveira DB, Balvedi MCW, da Silva RA, Aguiar AF. L-Arg supplementation does not improve muscle function during recovery from resistance exercise. Appl Physiol Nutr Metab 2018; 43:928-936

55. Birol A, Nişancı Kılınç F, Deliceoğlu G, Keskin ED. The effect of acute L-Arg supplementation on repeated sprint ability performance. Progr Nutr 2019; 21 :5-11.
56. Bode-Böger SM, Böger RH, Galland A, Tsikas D, Frölich JC. L-Arg-induced vasodilation in healthy humans: pharmacokinetic-pharmacodynamic relationship. Br J Clin Pharmacol 1998; 46: 489-497
57. Bode-Böger SM, Muke J, Surdacki A, Brabant G, Böger RH, Frölich JC. Oral L-Arg improves endothelial function in healthy individuals older than 70 years. Vasc Med 2003; 8:77-81
58. Fahs CA, Heffernan KS, Fernhall B. Hemodynamic and vascular response to resistance exercise with L-Arg. Med Sci Sports Exerc 2009; 41 :773-779
59. Forbes SC, Bell GJ. The acute effects of a low and high dose of oral L-Arg supplementation in young active males at rest. Appl Physiol Nutr Metab 2011; 36 :405-411
60. Forbes SC, Harber V, Bell GJ. The acute effects of L-Arg on hormonal and metabolic responses during submaximal exercise in trained cyclists. Int J Sport Nutr Exerc Metab 2013; 23 :369-377
61. Mansoor JK, Morrissey BM, Walby WF, Yoneda KY, Juarez M, Kajekar R, Severinghaus JW, Eldridge MW, Schelegle ES. L-Arg supplementation enhances exhaled NO, breath condensate VEGF, and headache at 4,342 m. High Alt Med Biol 2005; 6:289-300
62. Meirelles CM, Matsuura C. Acute supplementation of L-Arg affects neither strength performance nor nitric oxide production. J Sports Med Phys Fitness 2018; 58: 216-220
63. Robinson TM, Sewell D, Greenhaff PL. L-Arg ingestion after rest and exercise: effects on glucose disposal. Med Sci Sports Exerc 2003; 35:1309-1315
64. Schwedhelm E, Maas R, Freese R, Jung D, Lukacs Z, Jambrecina A, Spickler W, Schulze F, Böger RH. Pharmacokinetic and pharmacodynamic properties of oral L-citrulline and L-Arg: impact on nitric oxide metabolism. Br J Clin Pharmacol 2008; 65 :51-59
65. Suzuki T, Morita M, Hayashi T, Kamimura A. The effects on plasma L-Arg levels of combined oral L-citrulline and L-Arg supplementation in healthy males. Biosci Biotechnol Biochem 2017; 81 :372-375
66. Vanhatalo A, Bailey SJ, DiMenna FJ, Blackwell JR, Wallis GA, Jones AM. No effect of acute L-Arg supplementation on O_2 cost or exercise tolerance. Eur J Appl Physiol 2013; 113 :1805-1819
67. Vuletic L, Spalj S, Rogic D, Ruzic L, Alajbeg I. Effect of L-Arg dietary supplementation on salivary urea concentration and pH in physically active individuals. Aust Dent J. 2013; 58:491-497
68. Holeček M. Side effects of amino acid supplements. Physiol Res. 2022; Mar 25 ;71 :29-45
69. Barbul A. Arginine: biochemistry, physiology, and therapeutic implications. JPEN J Parenter Enteral Nutr 1986; 10:227-238

70. Evans RW, Fernstrom JD, Thompson J, Morris SM Jr, Kuller LH. Biochemical responses of healthy subjects during dietary supplementation with L-arginine. J Nutr Biochem 2004; 15:534-539
71. Hathcock JN, Shao A. Expanded approach to tolerable upper intake guidelines for nutrients and bioactive substances. *J Nutr* 2008; 138: 1992S-1995S

Chapter 4.

CONCLUDING REMARKS

We developed a model that accounts for the heterogeneity between studies in the safety assessment of amino acids through systematic review.
In this study, we proposed the w-CPRM as part of the solution and confirmed the necessity of utilizing heterogeneity among studies as weights for threshold estimation.

Currently, many amino acids have been used as functional foods and dietary supplements. However, there are no guidelines concerning their Tolerable Upper Intake Levels (ULs) for safe use. In general, Acceptable Daily Intakes (ADIs) for chemicals are calculated from NOAEL in animal studies using safety factor (SF) of 100. However, this general approach cannot be applied to determine ULs for amino acids as is discussed inn Chapter 1. Therefore, the data obtained from the human study is important for the safety assessment of amino acids.
In recent years, a systematic review (SR) has been used as an approach for assessing safety as well as effectiveness. We proposed a method to estimate the NOAEL and LOAEL using SR and reported them of L-lysine, one of the amino acids. However, these values we have reported are nearly the same as the value called the observed safety level (OSL) and it remains a problem to decide the way to estimate the threshold dose (TD). We believe that TD corresponds to the point of adverse events (AEs) occurrence, and calculating TD is important in deciding UL.

Therefore, we decided to explore the threshold by the Change-Point Regression Model （CPRM） by applying the hockey stick model, recently used in toxicological studies. The model is characterized by a biphasic manner, with regions that are not dose-responsive and regions that are dose-responsive. Theoretically, the TD should be between NOAEL and LOAEL.

By the way, in an SR with a meta-analysis (MA), either a fixed or random-effects model is applied, conducting a weighted synthetic analysis for each study. Recognizing the importance of incorporating weights for each study, we extended this approach to the CPRM. Our application focused on a clinical study targeting L-arginine (L-Arg) since L-lysine had limited intervention trials with comparisons to a control group, such as randomized control trials. Gastrointestinal symptoms were the most frequently reported as adverse events, therefore, we conducted by focusing on the occurrence of these events. We also decided that it is crucial to consider the use

of a one-time dose because gastrointestinal symptoms are considered a relatively acute response. Therefore, we considered L-Arg dose per day and per one-time, respectively.

【w-CPRM】
w-CPRM is indicated,

$$\omega_{1i} y_i = \omega_{1i} \beta \mathrm{I}(\chi_i > \chi_{cp})(\chi_i - \chi_{cp}) + \omega_{1i} \varepsilon_i$$

where y_i is the risk difference of the probability of reaction of L-Arg and placebo group and x_i is the dosage of L-Arg for the *i*-th study. However, x_{cp} is the unknown change point or threshold. I(・) is an indicator function defined as $I(x_i > x_{cp}) = 1$ if $x_i > x_{cp}$ and is 0 otherwise. ω_{1i} is the weight for each study.

We used the result of SR to assess the prevention of cardiovascular disease by omega-3 fatty acids as application data and validated the necessity of adjusting by weights in CPRM. From the result, the fitting of the model was better and we confirmed the necessity of weights.

Incidentally, on the target study of SR, we have conducted a risk of bias (RoB) assessment by using the way recommended by Cochrane Community and Jadad score.

Despite narrowing the study design to the highest quality RCTs, the trials collected for the safety evaluation of L-Arg were not homogeneous, with some variation in RoB ratings between studies. Therefore, we used the results of the RoB assessment recommended by The Cochrane Collaboration which is widely used in SR, and set the weight to $\omega_{2i.}$

【TD of L-Arg】
In this study, we focused on 23 studies reporting information on one-time doses out of the 24 studies reporting the presence or absence of AEs.

The RoB assessment of the Cochrane Collaboration consists of all 7 items and is assessed in three stages: low risk, high risk and unclear. we scored these results as “low risk” = 1, “high risk” = -1, and "unclear" = 0 for each risk of bias item and these full scores are 7 and are called $score_i$.
We used this $score_i$ and defined ω_{2i} as follows.

$$\omega_{2i} = \omega_{1i} + RS_i = \frac{1}{SE_i^2 + \widehat{\tau^2}} + RS_i, \quad RS_i = \frac{score_i}{(\sum_{i=1}^{n} score_i)}$$

On the other hand, Morizane has proposed a method to adjust bias in meta-analysis using Chalmers's score, an evaluation approach for Randomized Control Trials (RCTs) that is not commonly employed. Therefore, we utilized the Risk of Bias (RoB) tool from The Cochrane Collaboration, which is currently generally used. We considered the method of calculating weights based on Morizane's revised approach as follows.

$$\omega_{3i} = \omega_{1i} * RS_i = \frac{1}{SE_i^2 + \widehat{\tau^2}} * RS_i, \quad RS_i = \frac{score_i}{14}$$

Here, for ω_{3i}, the total value calculated as 'low risk' = +2, 'unclear' = +1, and 'high risk' = 0 was used as $score_i$.

Both of ω_{2i} and ω_{3i} used the results of the same RoB assessments but the results were significantly different. The threshold dose for ω_{3i} is 8754 mg/one-time dose, a lower dose than ω_{2i}. We confirmed that the model for ω_{3i} fits better than ω_{2i}.

The threshold dose of L-Arg was conventionally estimated at 12000 mg/day (4000 mg × 3 times) using the OSL. However, from the AIC profile of the w-CPRM using ω_1 as a weight, the AIC was -52.83 for 4000 mg/one-time dose. The AIC improved to -62.31 for 7531 mg/one-time dose, suggesting that this value is closer to the true threshold dose than the estimate obtained from OSL. The result using ω_{1i}, which is commonly used as weight in MA has the smallest AIC compared to other two weights used this time and the fitting of the model was best. Therefore, we judged the threshold dose of digestive symptoms in L-Arg intake targeted for healthy human is 7531 mg/one-time dose.

The result using ω_{1i}, commonly employed as a weight in an MA, yielded the smallest AIC compared to the other two weights used in this analysis, indicating the best model fit.
Consequently, we determined the threshold dose for gastrointestinal symptoms in L-Arg intake targeted for healthy humans to be 7531 mg/one-time dose (Fig.1).

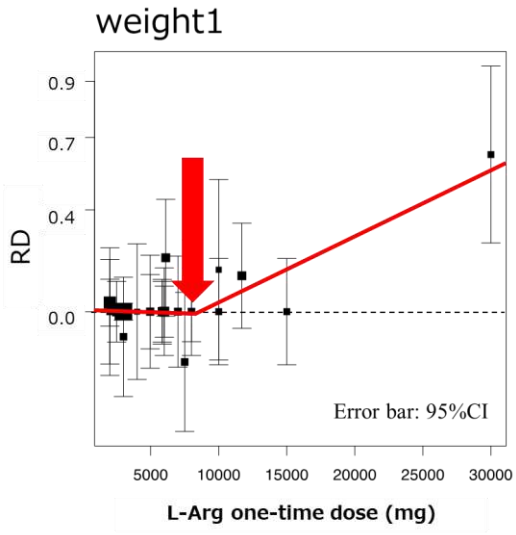

Fig. 1 Result of w-CPRM using weight1 for L-Arg study.
The x- and y-axes indicate the dosage of L-Arg and the RD of the occurrence of digestive symptoms, respectively. The size of black squares indicates the size of weights of each study and error bars are drawn above and below them. The threshold dose is 7531 mg/one-time dose.
RD: risk difference.

It is impossible to obtain this TD using the conventional OSL-based method. This issue was solved by using w-CPRM. By the way, the estimated TD was the point estimation value in this study. Therefore, we calculated the variation in the estimated TD of L-Arg when using ω_1, which was ± 1260 mg/one-time dose. Since the purpose of this study is to evaluate safety, it is important to carefully present the estimated TD. Therefore, it is important to consider the information on the lower limit of the estimated TD.

We applied ω_2 and ω_3 were scored based on the results of the RoB assessment and used in addition to ω_1. Several meta-analysis methods adjusted for RoB have been proposed. However, they are not widely used due to the limited number of reports [1,2]. Meanwhile, RoB assessment is a specific challenge in systematic reviews, reflecting some of the heterogeneity of bias risk between trials. Although we reported the results using the most fitting ω_1, if conducting TD estimation using a systematic review, it is also necessary to incorporate the results of RoB assessment as weighting factors.

However, the estimated TD was significantly changed when the RoB was incorporated into the model as weights. In addition, although we used the RoB assessment of the Cochrane Collaboration which is widely and generally used this time, it is important to note that the choice of RoB assessment method could potentially impact the estimated TD.

In this study, we reported the TD as one-time dose on the occurrence of gastrointestinal symptoms of L-Arg. From the result of the sensitivity analysis, we also have confirmed the statistical significance of the dose dependency regarding one-time dose. However, since the safety of chemicals is generally evaluated using ADI, it is ultimately necessary to report the TD as a daily dose. Additionally, if it is to be set as an upper limit, consideration should be given to the average intake from the diet.

The application of the w-CPRM in systematic reviews is not limited to the safety evaluation of amino acids but may also be useful in establishing ADIs and tolerable upper limits, especially for food ingredients with fewer tolerability studies. However, many issues remain in the use of RoB evaluation considering heterogeneity between studies, and further research is needed to resolve these issues.

REFERENCES

1. Turner RM, Spiegelhalter DJ, Smith GC, Thompson SG. Bias modelling in evidence synthesis. J R Stat Soc Ser A Stat Soc. 2009 Jan; 172 (1): 21-47.
2. Chalmers TC, Smith H Jr, Blackburn B, Silverman B, Schroeder B, Reitman D, Ambroz A. A method for assessing the quality of a randomized control trial. Control Clin Trials. 1981 May; 2(1): 31-49.

ACKNOWLEDGEMENTS

Safety assessment of L-Arg oral intake in healthy subject: a systematic review of randomized control trials (Chapter 3.) was supported by the Research Program on the Development of Innovative Technology Grants from the Project of the Bio-oriented Technology Research Advancement Institution (BRAIN: JPJ007097-04013B2).

This study was performed at the Laboratory of Food Chemistry, Yokohama University of Pharmacy.
I am most grateful to Dr. Kohsuke Hayamizu, Professor of Yokohama University of Pharmacy, for his kind guidance, invaluable discussions, and critical reading of this manuscript.
I am also greatly indebted to Dr. Ryosei Sakai, Visiting Professor of Yokohama University of Pharmacy, for his suggestions regarding whole background of amino acids and the conduction of w-CPRM.
I would also like to express my thanks to Dr. Nobumitsu Hanioka, Professor of Yokohama University of Pharmacy, for his helpful suggestions from the wide chemical aspect.
I would also like to thank Dr. Hideko Sone, Professor of Yokohama University of Pharmacy, for invaluable discussions from the aspect of toxicology.
I would also like to thank Dr. Akihide Sumino, Senior Lecturer of Yokohama University of Pharmacy, for his constant encouragement.
My heartfelt thanks are also due to the past and present members of the Laboratory of Food Chemistry, Yokohama University of Pharmacy, for their helpful cooperation and kind assistance.
In particular, I would like to thank Ms. Mai Murata for conducting data extraction and management and assessing the RoB on the safety assessment of L-Arg together.

APPENDIX

Table S1 Leave-one-out sensitivity and heterogeneity analyses for L-Arg

	L-Arg (n)	Placebo (n)	Quantitative data synthesis				Heterogeneity analysis			
			overall effect size	95%Cl	Z value	P value	Tau^2	Chi^2	df(Q)	I^2
Overall effect	314	333	0.01	-0.02 : 0.05	0.87	0.39	0.00	24.81	22	11%
Leave-one-out sensitivity analysis										
Ast et al. 2011	307	327	0.02	-0.02 : 0.05	0.89	0.37	0.00	24.95	21	16%
Pahlavani et al. 2014	286	305	0.01	-0.02 : 0.05	0.73	0.47	0.00	24.68	21	15%
Suzuki et al. 2017	304	301	0.02	-0.02 : 0.05	0.91	0.36	0.00	25.06	21	16%
Ueno et al. 2018	299	318	0.02	-0.02 : 0.05	0.92	0.36	0.00	25.01	21	16%
Alvares et al. 2012	306	326	0.02	-0.02 : 0.05	0.89	0.37	0.00	24.96	21	16%
Ast et al. 2011 (2)	308	327	0.02	-0.02 : 0.05	0.89	0.37	0.00	24.94	21	16%
Blum et al. 2000	304	323	0.02	-0.02 : 0.05	1.00	0.32	0.00	24.97	21	16%
Forbes and Bell 2011	300	319	0.02	-0.02 : 0.05	0.91	0.36	0.00	25.01	21	16%
Forbes et al. 2013	299	318	0.02	-0.02 : 0.05	0.92	0.36	0.00	25.01	21	16%
Meirelles and Matsuura 2018	302	321	0.02	-0.02 : 0.05	0.91	0.36	0.00	25.01	21	16%
Vuletic et al. 2013	255	275	0.02	-0.02 : 0.05	1.04	0.3	0.00	21.98	21	4%
Adams et al. 1995	302	321	0.02	-0.02 : 0.05	0.90	0.37	0.00	25.11	21	16%
Aguiar et al. 2016	304	323	0.02	-0.02 : 0.05	0.90	0.37	0.00	24.99	21	16%
Andrade et al. 2018	304	323	0.02	-0.02 : 0.05	0.90	0.37	0.00	24.99	21	16%
Bode-Boger et al. 2003	302	321	0.01	-0.02 : 0.04	0.76	0.45	0.00	23.95	21	12%
Forbes et al. 2014	300	319	0.01	-0.02 : 0.03	0.54	0.59	0.00	20.14	21	0%
Luiking et al. 1998	304	323	0.02	-0.02 : 0.05	1.05	0.29	0.00	24.59	21	15%
Vanhatalo et al. 2013	296	315	0.02	-0.02 : 0.05	0.93	0.35	0.00	24.99	21	16%
Chin-Dusting et al. 1996	306	325	0.02	-0.02 : 0.05	0.90	0.37	0.00	24.97	21	16%
Forbes and Bell 2011 (2)	300	319	0.01	-0.02 : 0.04	0.65	0.52	0.00	22.24	21	6%
Robinson et al. 2003	308	327	0.01	-0.02 : 0.04	0.79	0.43	0.00	23.50	21	11%
Savoye et al. 2006	306	325	0.02	-0.02 : 0.05	0.90	0.37	0.00	24.97	21	16%
Savoye et al. 2006 (2)	306	325	0.01	-0.02 : 0.03	0.48	0.63	0.00	10.24	21	0%

Table S2 Details for adverse events

Study	Adverse events		Daily dose (mg)	One-time dose (mg)	Duration of trial (day)	Category
	L-Arg	Placebo				
Adams et al. 1995	Abdominal bloating (n=1) Mild headaches (n=1)	Abdominal bloating (n=1)	21000	7000	3	C
Blum et al. 2000	Bullous pemphigoid (n=1: drop out)	Abdominal pain (n=1: drop out)	9000	3000	30	B
Bode-Böger et al. 2003	Diarrhea (n=1)	-	16000	8000	14	C
Forbes and Bell 2011	Mild GI distress (n=2)	-	11700	11700	1	C
Forbes et al. 2014	Vomited (n=3: drop out) Light headache (n=1: drop out)	-	6150	6150	1	C
Luiking et al. 1998	-	Nausea (n=2)	30000	7500	8	B
Pahlavani et al. 2014	Stomach problem (n=:1drop out) Skin dermatitis (n=2: drop out)	-	2000	2000	45	C
Robinson et al. 2003	Stomach discomfort (n=1)	-	10000	10000	1	C
Savoye et al. 2006	Bloating and diarrhea (n=5)	-	30000	30000	1	C

Table S3 Search terms

Web databases	**Search terms**
PubMed	L-Arginine [Ti] +double-blind(filters)
Cochrane Library	L-Arginine [Ti]+ double-blind
Ichushi-Web	L-Arginine [Ti] or アルギニン [Ti] (原著＋ランダム＋ヒト)
EBSCOhost	L-Arginine [Ti]+double-blind +randomized controlled trials