Random serial sampling to evaluate efficacy of iron fortification: a randomized controlled trial of margarine fortification with ferric pyrophosphate or sodium iron edetate

Maria Andersson, Winfried Theis, Michael B Zimmermann, Jasmin Tajeri Foman, Martin Jäkel, Guus SMJE Duchateau, Leon GJ Frenken, and Richard F Hurrell

ABSTRACT

Background: Random serial sampling is widely used in population pharmacokinetic studies and may have advantages compared with conventional fixed time-point evaluation of iron fortification.

Objective: Our objective was to validate random serial sampling to judge the efficacy of iron fortification of a low-fat margarine.

Design: We conducted a 32-wk placebo-controlled, double-blind, iron-intervention trial in 18–40-y-old Swiss women (n = 142) with serum ferritin (SF) concentrations <25 µg/L. Women were randomly assigned to 3 groups to receive 20 g margarine, with 14 mg added iron as either micronized ground ferric pyrophosphate (MGFePP) or sodium iron edetate (NaFeEDTA), or placebo daily. We measured hemoglobin and iron status of subjects at 2 fixed time points (at baseline and the endpoint) plus 3 randomly assigned time points between 4 and 28 wk. With the use of bootstrapping, the number of observations per individual was reduced to 3 and then compared with the 5-time-point data. Mixed-effects models were used to estimate iron repletion over time for random sampling, and analysis of covariance was used for fixed time-point sampling.

Results: Body iron stores increased in women who received MGFePP or NaFeEDTA compared with women who received placebo (P < 0.05). The increase in body iron stores with NaFeEDTA fortification was 2–3 times the increase with MGFePP fortification (P < 0.05); the difference was more marked in women with baseline SF concentrations <15 µg/L (P < 0.05). Random serial sampling reduced the required sample size per group to one-tenth of that for 2 fixed time points. Compared with the 5-time-point analysis, the 3-time-point sparse sampling generated comparable estimates of efficacy.

Conclusions: When used to evaluate the efficacy of iron fortificants, random serial sampling can reduce the sample size, invasiveness, and costs while increasing sensitivity. Random serial sampling more clearly describes the pattern of iron repletion and may prove useful in evaluating other micronutrient interventions.

INTRODUCTION

Iron fortification of foods can be a cost-effective, long-term, population-based strategy to improve iron status and to prevent iron deficiency worldwide (1, 2). An array of iron fortificants suitable for different food vehicles is available (3). However, the extent of the effect of iron-fortified foods on iron status in the population depends on several factors, such as the amount of iron lacking in the diet, the amount of fortification iron, the bioavailability of the iron fortificant, the food-matrix, the frequency of consumption of the fortified food, the iron status of the individual, and the overall nutritional status of the target population (1, 4). The efficacy of iron-fortification strategies is generally evaluated by longitudinal, population-based, randomized-controlled trials carried out over a time frame of 6–9 mo with ≥2 fixed time points for follow-up (1, 3, 5). Iron fortification is considered efficacious when it significantly improves iron status, reduces the prevalence of iron deficiency in a population by ≥10% (6), or the prevalence in the population is reduced to <10% (7). However, conducting an efficacy trial is costly, invasive, and logistically demanding.

In recent years new methods of clinical study design and data analysis have become available (8, 9). These advanced methods can increase the efficiency, probability of success, and cost effectiveness of clinical trials (10). Longitudinal random serial sampling is one method widely used in population pharmacokinetic studies (11–14). Random sampling design typically involves blood collection at one or more randomly allocated time points from each subject that are uniformly distributed over the study period. A sparse random-sampling strategy is commonly applied to reduce the number of blood samples from each individual (13, 15), and mixed-effects models (MEMs) for repeated measurements are used to assess the population mean (9). This approach increases the power of the study and may have advantages over conventional fixed time-point evaluation in population-based efficacy evaluations of micronutrient interventions.

1 From the Human Nutrition Laboratory, Institute of Food, Nutrition, and Health, Swiss Federal Institute of Technology Zürich, Zurich, Switzerland (MA, JTF, MBZ, and RFH); the Human Nutrition Division, Wageningen University, Wageningen, Netherlands (MBZ); and Unilever Research and Development, Vlaardingen, Netherlands (WT, GSMJED, MJ, and LGJF).

2 Supported by Unilever Research and Development (Vlaardingen, Netherlands) and the Swiss Federal Institute of Technology (Zürich, Switzerland). Unilever Research and Development provided the margarine.

3 Address correspondence to M Andersson, Human Nutrition Laboratory, Institute of Food, Nutrition and Health, Swiss Federal Institute of Technology Zürich, Schmelzbergstrasse 7, CH-8092 Zurich, Switzerland. E-mail: maria.andersson@itw.agrif.ethz.ch.

Received March 16, 2010. Accepted for publication July 30, 2010.
First published online September 15, 2010; doi: 10.3945/ajcn.2010.29523.
Therefore, the aim of this study was to develop and validate a longitudinal sampling approach that uses a random serial-sampling methodology for evaluating the efficacy of iron-fortified foods in populations. We evaluated margarine as a new vehicle for iron fortification in industrialized countries and measured the efficacy of iron fortification with micronized ground ferric pyrophosphate (MGFePP) or sodium iron edetate (NaFeEDTA) in women of reproductive age with low iron stores. We compared the new efficacy study design based on random serial sampling or sparse random serial sampling with a conventional design based on 2 time points at the start and the end of the study.

SUBJECTS AND METHODS

Fortification of low-fat margarine

A vegetable oil–based low-fat margarine was developed by Unilever Research and Development (Vlaardingen, Netherlands) and used as vehicle for the iron fortification in this study. The nutrient composition was similar to current commercially available low-fat margarines with a total fat content of 40% and 60% water. The trans fatty acid content of the fat component was <1%. The low-fat margarine was fortified with MGFePP (24.6% Fe, mean particle size: 2.5 μm; Paul Lohmann AG, Emmerthal, Germany) or with NaFeEDTA (13.6% Fe; Akzo Nobel Chemicals Pte Ltd, Singapore) at an amount of 0.70 g Fe/kg margarine. The margarine was packed into individual 20-g daily portions. The 20 g amount was chosen to match the portion size for the reference labeling values (RLVs) of the European Union (16) and the average spreadable margarine and butter intake in women in northern and central Europe (17, 18). The daily 20 g serving in the 3 groups contained no added iron (group 1), 14 mg Fe as MGFePP (group 2), and 14 mg Fe as NaFeEDTA (group 3). The fortification amount was based on the RLV in Europe for iron (14 mg), and it supplied 100% of the RLV per daily 20-g portion (19, 20), which corresponded to 80% of the recommended dietary allowance for iron in women of reproductive age (21).

The low-fat margarine underwent sensory, acceptability, and microbiological testing before study start. The shelf life and storage stability of the margarine used in the current study was 3 mo when stored at 4°C. The margarine was produced and distributed monthly in 10 different production batches. The mean iron (±SD) content in the 10 batches was 0.711 ± 0.011 g Fe/kg and 0.709 ± 0.015 g Fe/kg for MGFePP and NaFeEDTA, respectively.

Study subjects

The study was conducted in healthy young women in the Zurich area of Switzerland (Figure 1). Women aged 18–40 y old were recruited locally through advertisements at colleges and universities and invited to a screening of iron deficiency. Blood was drawn from volunteers, and serum ferritin (SF) concentrations were measured. The inclusion criteria for the intervention trial were as follows: 1) female gender, 2) body mass index (BMI; in kg/m²) of 17.0–28.5, 3) SF concentration <25 μg/L, with or without anemia, 4) unelevated C-reactive protein (CRP) concentration (<10 mg/L), 5) not pregnant or breastfeeding, 6) no blood donation during the 4 mo preceding the study, and 7) no use of chronic medications (except hormonal contraceptives; 57% of women were using hormonal contraception). No
vitamin or mineral supplements were allowed during the study and vitamin and mineral supplement use was discontinued 2 wk before the start of the study. The study protocol was approved by the Ethical Committee of the Swiss Federal Institute of Technology Zürich (Zurich, Switzerland). All participants gave informed consent to participate in the study. The study was conducted according to the Declaration of Helsinki (22). The study period was September 2007 to July 2008.

Iron intervention

The study was a randomized, double-blind, controlled intervention trial in women aged 18–40 y with low iron stores. Of the 673 women who were screened for iron deficiency, 263 were eligible and invited to participate in the study. Thirty women declined participation before the start of the study, and 91 women were enrolled in additional arms not reported. Thus, 142 women were enrolled and randomly assigned to receive 20 g low-fat margarine/d that contained 1) no added iron, 2) 14 mg Fe as MgFePP, or 3) 14 mg Fe as NaFeEDTA. Enrollment was staggered, and subjects were enrolled at 4 different time points over 2 mo. Investigators and study participants were blinded to group assignment. Codes were revealed only after all subjects had completed the trial and after a blind review of the data.

The low-fat margarine was provided to study participants monthly for 32 wk in 30 daily portion packs (20 g each) at individual visits to the study center at the Swiss Federal Institute of Technology Zürich. At the start of the study, it was emphasized that the study margarine should be consumed as spread on bread and pastries, preferably for breakfast. Subjects were instructed to consume any remainder of the daily portion as margarine on bread for lunch or dinner. Because of the instability of low-fat margarine during heating, study participants were instructed not to use the low-fat margarine for cooking, in food preparation, or on hot food. The importance of the complete consumption of the provided daily portion was emphasized at the start of the study and was reinforced at each monthly visit. For monitoring, all used margarine containers were collected monthly, and the weight of the returned portion packages was recorded. The individual compliance of margarine consumption was estimated by calculating the proportional difference between the daily consumed and intended servings.

Data collection

Random serial sampling

A total of 5 serial blood samples were collected from each study participant over the 32-wk study period. The first and last sampling time points were fixed, and each subject was measured at baseline and at the end of the study at 32 wk. In between these time points, 3 sample points were randomly allocated within 3 sampling blocks. The intermediate blood samples were drawn at randomly generated time points from each of 3 9-wk blocks between the start and end of the study. Each sample was collected a minimum of 2 wk apart at ≥4 wk from the start of the study and 2 wk from the end of the study. The gap of ≥4 wk at the start of the study was chosen to allow for a change in iron status to take place after the initiation of the intervention. The sampling dates of the intermediate samples were generated in such a way that data points from all subjects within each group were uniformly distributed over the 32 wk to allow curve fitting. Sampling was conducted on the same 2 weekdays every week. At each visit, subjects were weighed and a questionnaire on general health was completed. At the end of the study, women who remained iron deficient were referred to their personal physician for treatment.

Sparse random serial sampling

A sparse random serial-sampling design was simulated by creating a subset of data drawn from the complete longitudinal serial-sampling data set. A bootstrap procedure was used to reduce 5 observations per individual to a sparse serial-sampling scheme with 3 observations per subject (23). All subjects who provided ≥3 blood samples were used for the bootstrapping procedure to resample 3 randomly selected observations for each subject. The sparse-sampling design model was obtained by repeatedly fitting 100 bootstrap samples from the longitudinal serial-sampling data set for each study group.

Laboratory analyses

The determination of the iron content in the margarine was measured by using inductively coupled plasma atomic-emission spectrometry (Perkin Elmer 3300 DV; Perkin Elmer, Waltham, MA). The samples were digested in nitric acid and hydrogen peroxide in closed vessels in a microwave oven (CEM Mars 5; CEM, Matthews, NC) at 200°C and pressure (maximum: 110 bar). Samples were measured at 238.204 nm.

Blood samples were analyzed for hemoglobin, SF, transferrin receptor (TfR), and CRP concentrations. Hemoglobin was measured in whole blood on the day of collection with an automated Coulter counter (AcT8 Counter; Beckman Coulter, Krefeld, Germany). Blood was centrifuged (3000 rpm for 15 min at room temperature), and serum was separated into aliquots (0.5 mL) and frozen at −25°C until analysis. SF and CRP were measured on an IMMULITE automatic system (DPC Bühlmann GmbH, Aschwill, Switzerland). External 3-level commercial quality-control materials were used for the measurement of hemoglobin, SF, and CRP (Beckman Coulter; World Health Organization Standard 80/578, Ramco Laboratories Inc, Houston, TX; Digitana AG, Horgen, Switzerland; and DPC Bühlmann GmbH). TfR was measured with an automated immunoturbidimetric assay (Roche Diagnostics, Mannheim, Germany) (24, 25). Anemia and iron deficiency were defined as a hemoglobin concentration <12 g/dL and an SF concentration <15 μg/L, respectively (26). Iron-deficiency anemia was defined as concurrent anemia and iron deficiency. The manufacturer-specified normal reference range for adult premenopausal women for TfR concentrations measured by the Roche method (Roche Diagnostics) is 1.9–4.4 mg/L (24). A TfR concentration >4.4 mg/L indicated elevated TfR and CRP concentration ≥10 mg/L, suggested the presence of inflammation. Data points with SF >15 μg/L from subjects with elevated CRP concentrations ≥10 mg/L (29 out of 984 data points) were excluded from the analyses of SF and body iron.

Body iron stores (in mg Fe/kg body weight) was calculated for each subject from the TfR:SF ratio by using the algorithm by Cook et al (27, 28); for this calculation we converted TfR concentrations by using the correction factor suggested by Pfeiffer.
et al (25). Depleted body iron stores were defined as body iron stores <0 mg Fe/kg body weight (27, 28). Iron absorption during the study in the 2 iron groups, compared with that of the placebo group, was calculated by comparing the total amount of fortification iron consumed with the mean change in body iron stores.

Statistical analyses

The study was designed to investigate the efficacy of MGIFePP and NaFeEDTA fortification compared with that of a placebo by using MEMs for repeated measurements. MEMs estimate the overall mean population response over time and take both population variability (fixed effects) and individual variability (random effects) into account in the analyses (9, 12, 29, 30).

Sample-size calculations assumed a linear association for MEMs over time. At the start of the study, we anticipated a ≥15% change from a baseline SF concentration of 15 µg/L in the iron groups compared with that in the placebo group over 16 wk (half-time). The sample sizes required for slope testing of theoretical regression lines to attain a 90% power at a significance level of 5% were 68 subjects for MGIFePP and 10 subjects for NaFeEDTA, respectively. Thus, a minimum of 34 women per group were required. To account for dropouts and to allow for data fitting for each group separately, ≥47 women per group were included.

Descriptive statistics are reported as means ± SDs for normally distributed data, geometric means ± SDs for log-transformed data, or medians (ranges) for data that remained nonnormal after log transformation. Group differences at baseline for continuous variables were tested by using analysis of variance and Student’s t test for normally distributed data and the Kruskal-Wallis test and Mann-Whitney rank sum test for non-normally distributed data. Group differences at baseline for categorical binary variables were tested by using Pearson’s chi-square test. Statistical analyses were by intent-to-treat analyses for subjects with one or more follow-up measurements in addition to baseline. One subject with elevated CRP at 4 time points (in the NaFeEDTA group) was excluded from the data analysis.

MEMs for repeated measurements were derived separately for each outcome variable (hemoglobin, SF, TfR, and body iron) with time (days of follow-up) and dose (iron intake) interactions as fixed effects for each study group. The random effects (individual variations) were introduced for the intercept and slope of every subject’s set of repeated measurements. The normality of the residuals and the homogeneity of variance were examined by plotting and residual plots for each model. Data with nonnormal distribution of residuals after regression were log transformed. Plots and correlation coefficients were used to verify the fit. The models assumed a proportional-to-dose effect (ie, the response in outcome variables was assumed to be proportional to the amount of iron consumed over the 32 wk). Different proportionality-to-dose factors were assumed for MGIFePP and NaFeEDTA. Two fixed effects for dose were tested as follows: dose as the intention-to-treat dose and dose as the actual consumed dose (compliance corrected). The results were comparable, and only the analysis of actual consumed dose is presented. The efficacy of MGIFePP and NaFeEDTA fortification was tested by estimating the daily change in the iron-status variables compared with the placebo for a linear function. However, SF and body iron models showed a curved development, which we approximated by a quadratic function, that consisted of a linear and quadratic term. For SF and body iron, the linear term of the full quadratic model was compared with that of the placebo. The sparse-sampling design model was obtained by repeatedly fitting 100 bootstrap samples from the longitudinal serial-sampling data set for each study group. Categorical outcomes at 32 wk were modeled by binning the observations into months and subsequently applying a generalized linear model for binomial data with treatment group, month, and the group × month interaction as covariates.

The 3 different sampling schemes that were based on random serial sampling (5 samples), sparse serial sampling (3 samples), and conventional fixed time-point sampling design (2 samples at the beginning and end of the study) were compared by evaluating group differences in body iron stores over 32 wk obtained by a simple straight-line model for all 4 sampling strategies. Group differences and significance testing for the MEM estimates were also compared with conventional analysis of covariance (ANCOVA) with the baseline as the covariate for the fixed time-point data. P values are reported on a dimensionless scale to order the effect sizes and to compare across the 3 different sampling designs. The theoretical minimum sample size required for sparse random sampling was estimated as described and for fixed time-point evaluation by using ANCOVA with Bonferroni correction for multiple comparisons.

Descriptive statistics are reported as means ± SDs for normally distributed data, geometric means ± SDs for log-transformed data, or medians (ranges) for data that remained nonnormal after log transformation. Group differences at baseline for continuous variables were tested by using analysis of variance and Student’s t test for normally distributed data and the Kruskal-Wallis test and Mann-Whitney rank sum test for non-normally distributed data. Group differences at baseline for categorical binary variables were tested by using Pearson’s chi-square test. Statistical analyses were by intent-to-treat analyses for subjects with one or more follow-up measurements in addition to baseline. One subject with elevated CRP at 4 time points (in the NaFeEDTA group) was excluded from the data analysis.

MEMs for repeated measurements were derived separately for each outcome variable (hemoglobin, SF, TfR, and body iron) with time (days of follow-up) and dose (iron intake) interactions as fixed effects for each study group. The random effects (individual variations) were introduced for the intercept and slope of every subject’s set of repeated measurements. The normality of the residuals and the homogeneity of variance were examined by residual plots for each model. Data with nonnormal distribution of residuals after regression were log transformed. Plots and correlation coefficients were used to verify the fit. The models assumed a proportional-to-dose effect (ie, the response in outcome variables was assumed to be proportional to the amount of iron consumed over the 32 wk). Different proportionality-to-dose factors were assumed for MGIFePP and NaFeEDTA. Two fixed effects for dose were tested as follows: dose as the intention-to-treat dose and dose as the actual consumed dose (compliance corrected). The results were comparable, and only the analysis of actual consumed dose is presented. The efficacy of MGIFePP and NaFeEDTA fortification was tested by estimating the daily change in the iron-status variables compared with the placebo for a linear function. However, SF and body iron models showed a curved development, which we approximated by a quadratic function, that consisted of a linear and quadratic term. For SF and body iron, the linear term of the full quadratic model was compared with that of the placebo. The sparse-sampling design model was obtained by repeatedly fitting 100 bootstrap samples from the longitudinal serial-sampling data set for each study group. Categorical outcomes at 32 wk were modeled by binning the observations into months and subsequently applying a generalized linear model for binomial data with treatment group, month, and the group × month interaction as covariates.

The 3 different sampling schemes that were based on random serial sampling (5 samples), sparse serial sampling (3 samples), and conventional fixed time-point sampling design (2 samples at the beginning and end of the study) were compared by evaluating group differences in body iron stores over 32 wk obtained by a simple straight-line model for all 3 sampling strategies. Group differences and significance testing for the MEM estimates were also compared with conventional analysis of covariance (ANCOVA) with the baseline as the covariate for the fixed time-point data. P values are reported on a dimensionless scale to order the effect sizes and to compare across the 3 different sampling designs. The theoretical minimum sample size required for sparse random sampling was estimated as described and for fixed time-point evaluation by using ANCOVA with Bonferroni correction for multiple comparisons.

Descriptive statistics are reported as means ± SDs for normally distributed data, geometric means ± SDs for log-transformed data, or medians (ranges) for data that remained nonnormal after log transformation. Group differences at baseline for continuous variables were tested by using analysis of variance and Student’s t test for normally distributed data and the Kruskal-Wallis test and Mann-Whitney rank sum test for non-normally distributed data. Group differences at baseline for categorical binary variables were tested by using Pearson’s chi-square test. Statistical analyses were by intent-to-treat analyses for subjects with one or more follow-up measurements in addition to baseline. One subject with elevated CRP at 4 time points (in the NaFeEDTA group) was excluded from the data analysis.

MEMs for repeated measurements were derived separately for each outcome variable (hemoglobin, SF, TfR, and body iron) with time (days of follow-up) and dose (iron intake) interactions as fixed effects for each study group. The random effects (individual variations) were introduced for the intercept and slope of every subject’s set of repeated measurements. The normality of the residuals and the homogeneity of variance were examined by residual plots for each model. Data with nonnormal distribution of residuals after regression were log transformed. Plots and correlation coefficients were used to verify the fit. The models assumed a proportional-to-dose effect (ie, the response in outcome variables was assumed to be proportional to the amount of iron consumed over the 32 wk). Different proportionality-to-dose factors were assumed for MGIFePP and NaFeEDTA. Two fixed effects for dose were tested as follows: dose as the intention-to-treat dose and dose as the actual consumed dose (compliance corrected). The results were comparable, and only the analysis of actual consumed dose is presented. The efficacy of MGIFePP and NaFeEDTA fortification was tested by estimating the daily change in the iron-status variables compared with the placebo for a linear function. However, SF and body iron models showed a curved development, which we approximated by a quadratic function, that consisted of a linear and quadratic term. For SF and body iron, the linear term of the full quadratic model was compared with that of the placebo. The sparse-sampling design model was obtained by repeatedly fitting 100 bootstrap samples from the longitudinal serial-sampling data set for each study group. Categorical outcomes at 32 wk were modeled by binning the observations into months and subsequently applying a generalized linear model for binomial data with treatment group, month, and the group × month interaction as covariates.

The 3 different sampling schemes that were based on random serial sampling (5 samples), sparse serial sampling (3 samples), and conventional fixed time-point sampling design (2 samples at the beginning and end of the study) were compared by evaluating group differences in body iron stores over 32 wk obtained by a simple straight-line model for all 3 sampling strategies. Group differences and significance testing for the MEM estimates were also compared with conventional analysis of covariance (ANCOVA) with the baseline as the covariate for the fixed time-point data. P values are reported on a dimensionless scale to order the effect sizes and to compare across the 3 different sampling designs. The theoretical minimum sample size required for sparse random sampling was estimated as described and for fixed time-point evaluation by using ANCOVA with Bonferroni correction for multiple comparisons.

Data analyses were conducted with the R statistical programming environment with nlme, multcomp, and ggplot2 packages (R Development Core Team, 2007; http://www.R-project.org), SPSS (version 16.0; SPSS Inc, Chicago, IL), Excel Windows XP (Microsoft, Seattle, WA), and SAS (version 9.1; SAS Institute Inc, Cary, NC). Significance was set at P < 0.05.

RESULTS

Iron intervention

Anthropometric characteristics, hemoglobin, SF, and TfR concentrations, and body iron stores of women at baseline in the 3 groups are listed in Table 1. There were no significant differences between groups in any of the baseline characteristics. Although not significant, mean body iron stores were highest, and the number of individuals with depleted body iron stores was lowest in the NaFeEDTA group at baseline. Overall, anemia, iron deficiency, and iron deficiency anemia were present in 4.3%, 56.7%, and 4.3% of participating women, respectively, and inflammation (an elevated CRP) was present in 1.4% of participating women.

A total of 105 women completed the study, with an overall cumulative dropout rate of 13% of subjects at the second blood sample, 21% of subjects at the third blood sample, 25% of subjects at the fourth blood sample, and 26% of subjects at the fifth blood sample at 32 wk. Reasons for dropping out of the study and the respective frequency of the total 36 subjects who dropped out were as follows: margarine portion size was too large (n = 14), starting clinical iron treatment (n = 8), taste (n = 4), no response (n = 3), relocation (n = 1), medical condition (n = 2), no reason (n = 2), and pregnancy (n = 2). The dropout rate was not significantly different among treatment groups at any time point.
The mean daily amount of margarine consumed over the study period was 17.8 g or 89% of the supplied portion. Sixty-one percent of the women consumed 90% of the supplied portions, and compliance did not differ between treatment groups. Body weight and BMI did not change in any of the groups over the 32-wk study. There was no difference in the prevalence of elevated CRP values between the 3 intervention groups at any time point (data not shown).

**Overall effect of iron fortification determined by random serial sampling**

There was a strong correlation between the observed and predicted values estimated with MEM for all indicators. The correlations for body iron stores for the 3 intervention groups ($R^2 = 0.95, P < 0.001$) are shown in **Figure 2**. The effect of fortification iron as MGFePP and NaFeEDTA on hemoglobin, SF, and TIR concentrations and body iron stores over the 32-wk

### TABLE 1
Baseline characteristics of women after being randomly assigned to 3 intervention groups for the efficacy trial of iron-fortified, low-fat margarines

<table>
<thead>
<tr>
<th>Demographic characteristics</th>
<th>Placebo ($n = 48$)</th>
<th>MGFePP ($n = 47$)</th>
<th>NaFeEDTA ($n = 46$)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (y)</strong>$^2$</td>
<td>23.5 (19.2–32.1)</td>
<td>23.3 (18.5–40.7)</td>
<td>23.2 (19.0–40.9)</td>
</tr>
<tr>
<td><strong>Anthropometric measurements$^3$</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>$61.5 \pm 8.3$</td>
<td>$60.0 \pm 8.3$</td>
<td>$60.3 \pm 5.6$</td>
</tr>
<tr>
<td>Height (cm)</td>
<td>$168.3 \pm 7.0$</td>
<td>$166.8 \pm 7.0$</td>
<td>$168.3 \pm 7.2$</td>
</tr>
<tr>
<td>BMI (kg/m$^2$)</td>
<td>$21.7 \pm 2.4$</td>
<td>$21.5 \pm 2.4$</td>
<td>$21.3 \pm 2.0$</td>
</tr>
<tr>
<td><strong>Hemoglobin and iron status indicators</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hemoglobin (g/dL)$^4$</td>
<td>$13.5 \pm 0.9$</td>
<td>$13.7 \pm 0.9$</td>
<td>$13.6 \pm 0.9$</td>
</tr>
<tr>
<td>SF (μg/L)$^5,6$</td>
<td>$10.9 \pm 7.4$</td>
<td>$11.6 \pm 6.1$</td>
<td>$12.9 \pm 6.6$</td>
</tr>
<tr>
<td>TIR (mg/L)$^5$</td>
<td>$3.5 \pm 1.3$</td>
<td>$3.4 \pm 1.4$</td>
<td>$3.4 \pm 1.2$</td>
</tr>
<tr>
<td>Body iron (mg Fe/kg body weight)$^5,6$</td>
<td>$0.9 \pm 3.2$</td>
<td>$1.2 \pm 2.9$</td>
<td>$1.6 \pm 2.9$</td>
</tr>
<tr>
<td><strong>Prevalence [n (%)]</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anemia$^7$</td>
<td>1 (2.1)$^4$</td>
<td>2 (4.3)</td>
<td>3 (6.4)</td>
</tr>
<tr>
<td>Iron deficiency$^6,8$</td>
<td>29 (60.4)</td>
<td>28 (59.6)</td>
<td>23 (50.0)</td>
</tr>
<tr>
<td>Elevated TIR$^7$</td>
<td>12 (25.0)</td>
<td>10 (21.3)</td>
<td>9 (19.6)</td>
</tr>
<tr>
<td>Depleted body iron$^6,10$</td>
<td>18 (37.5)</td>
<td>13 (27.7)</td>
<td>11 (23.9)</td>
</tr>
</tbody>
</table>

$^1$ MGFePP, micronized ground ferric pyrophosphate; NaFeEDTA, sodium iron edetate; SF, serum ferritin; TIR, transferrin receptor. There were no significant differences in any of the baseline characteristics between the 3 groups. Weight, height, BMI, hemoglobin, body iron, log-transformed SF, and TIR were tested by using ANOVA and Student’s $t$ tests. Age was tested by using the Kruskal-Wallis test and Mann-Whitney rank sum test.

$^2$ Values are medians; ranges in parentheses.

$^3$ Values are means ± SDs.

$^4$ $n = 47$.

$^5$ Values are geometric means ± SDs.

$^6$ Samples with elevated C-reactive protein concentrations ($\geq 10$ mg/L) were excluded from the analysis.

$^7$ Defined as hemoglobin concentration $<12$ g/L.

$^8$ Defined as an SF concentration $<15$ μg/L.

$^9$ Defined as a TIR concentration $>4.4$ mg/L.

$^{10}$ Defined as a body iron store $<0$ mg Fe/kg body weight.

---

**FIGURE 2.** Correlation between observed and predicted values for body iron stores obtained with the mixed-effects regression model for repeated measurements for the 3 groups over 3 wk [placebo, $n = 48$; micronized ground ferric pyrophosphate (MGFePP), $n = 47$; sodium iron edetate (NaFeEDTA), $n = 46$] ($R^2 = 0.95, P < 0.001$).
### TABLE 2
Longitudinal effect of iron fortification on iron status in women in the 3 intervention groups over 32 wk

<table>
<thead>
<tr>
<th>Term</th>
<th>Placebo (n = 48)</th>
<th>MGFePP (n = 47)</th>
<th>NaFeEDTA (n = 46)</th>
<th>P²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept (95% CI)</td>
<td>n</td>
<td>Values</td>
<td>n</td>
<td>Values</td>
</tr>
<tr>
<td>Hemoglobin (g/L)</td>
<td>0</td>
<td>13.5 ± 0.9</td>
<td>0</td>
<td>13.7 ± 0.9</td>
</tr>
<tr>
<td>Week 32</td>
<td>0</td>
<td>13.3 ± 0.9</td>
<td>0</td>
<td>13.5 ± 0.7</td>
</tr>
<tr>
<td>SF (µg/L)</td>
<td>0</td>
<td>10.9 ± 7.4</td>
<td>0</td>
<td>11.6 ± 6.1</td>
</tr>
<tr>
<td>Week 32</td>
<td>0</td>
<td>11.9 ± 9.6</td>
<td>0</td>
<td>14.8 ± 11.8</td>
</tr>
<tr>
<td>TTR (mg/L)</td>
<td>0</td>
<td>3.5 ± 1.3</td>
<td>0</td>
<td>3.4 ± 1.4</td>
</tr>
<tr>
<td>Week 32</td>
<td>0</td>
<td>3.7 ± 1.8</td>
<td>0</td>
<td>3.2 ± 1.1</td>
</tr>
<tr>
<td>Body iron (mg/Fe/kg body weight)</td>
<td>0</td>
<td>0.9 ± 3.2</td>
<td>0</td>
<td>1.2 ± 2.9</td>
</tr>
<tr>
<td>Week 32</td>
<td>0</td>
<td>1.0 ± 3.8</td>
<td>0</td>
<td>2.2 ± 3.3</td>
</tr>
<tr>
<td>SF (µg/L)</td>
<td>0</td>
<td>0.00963 (0.0000493, 0.0000156)</td>
<td>0</td>
<td>0.00103 (0.000061, 0.0000015)</td>
</tr>
<tr>
<td>Week 32</td>
<td>0</td>
<td>0.00963 (0.0000493, 0.0000156)</td>
<td>0</td>
<td>0.00103 (0.000061, 0.0000015)</td>
</tr>
</tbody>
</table>

1 MGFePP, micronized ground ferric pyrophosphate; NaFeEDTA, sodium iron edetate; SF, serum ferritin; TTR, transferrin receptor. The table shows measured values at 0 and 32 wk plus estimated effects by using mixed-effects models (MEMs) for 5 repeated measurements over 0–32 wk. Models are presented as estimated values at baseline (intercept) and estimated daily changes in hemoglobin, SF, and TTR concentrations and body iron stores and their uncertainties predicted from the linear and quadratic MEM (SF and body iron) for repeated measures for the 3 intervention groups. The quadratic term provided a curvature in the longitudinal data, and values of daily changes in SF concentrations and body iron stores from the quadratic term as listed may have sped up or slowed down depending on the selected point in time (see Figure 3). No significant changes in hemoglobin, SF, and TTR concentrations and body iron stores were observed from baseline in the placebo group. Estimated daily changes in iron-status variables (ie, the slope of the regression line) for the MGFePP and the NaFeEDTA groups were compared with those in the placebo group.

2 P values indicate main effects compared with placebo.

3 Values are means ± SDs unless otherwise indicated.

4 Estimated daily change; 95% CI in parentheses (all such values).

5 Values are geometric means ± SDs unless otherwise indicated.

6 Estimates are shown as log-transformed data.
study period estimated from MEMs for repeated measurements is shown in Table 2. The estimated baseline value (intercept) and the estimated linear and quadratic terms (when appropriate) of the daily change and their uncertainties for repeated measures for each group separately are listed (Table 2). No significant changes in hemoglobin, SF, and TfR concentrations and body iron stores were observed from baseline in the placebo group. The estimated daily changes (linear term) in iron-status variables for the MGFePP and the NaFeEDTA groups are compared with those of the placebo group.

Mean hemoglobin concentration did not change in any of the study groups during the intervention (Table 2). Compared with placebo, iron fortification with MGFePP did not significantly improve SF or TfR concentrations, but it improved body iron stores ($P < 0.05$). Compared with placebo, NaFeEDTA significantly increased SF ($P < 0.001$), reduced TfR ($P < 0.001$) concentrations, and improved body iron stores ($P < 0.001$). Iron fortification with NaFeEDTA improved body iron stores to significantly higher amounts than did MGFePP ($P < 0.05$).

The population effect of iron fortification on body iron stores from the fitted fixed part of the MEM over 32 wk in the 3 study groups is shown in Figure 3. The thin lines in Figure 3 shows the individual development of body iron stores over time, and the thick line in Figure 3 indicates the average group development. The slope model predicted an average increase in body iron stores over 32 wk of 1.1 mg Fe/kg body weight (range: 0.3–1.2 mg Fe/kg body weight) for MGFePP and of 2.7 mg Fe/kg body weight (range: 2.1–3.5 mg Fe/kg body weight) for NaFeEDTA. This reflects the calculated absorption of the fortification iron during the study as follows: 2.8% from MGFePP and 5.9% from NaFeEDTA.

The effects of iron fortification on the prevalence of anemia, iron deficiency, elevated TfR concentrations, and depleted body iron stores for the 3 intervention groups are shown in Figure 4. The prevalence of anemia was 4.3% at baseline and did not change significantly in any of the 3 groups. Compared with the placebo, fortification with MGFePP did not significantly decrease the prevalence of iron deficiency ($P = 0.095$), but it significantly reduced the prevalence of elevated TfR concentrations ($P < 0.05$) and depleted iron stores ($P < 0.05$).

Compared with the placebo, fortification with NaFeEDTA reduced the prevalence of iron deficiency ($P < 0.001$), elevated TfR concentrations ($P < 0.001$), and depleted body iron status ($P < 0.001$). NaFeEDTA was more effective than MGFePP in reducing the prevalence of iron deficiency ($P < 0.001$), elevated TfR concentrations ($P < 0.05$), and the prevalence of depleted body iron stores ($P < 0.01$).

Effect of iron fortification on the basis of iron status at baseline

In a secondary analysis, we investigated the influence of iron status at baseline on the effect of iron fortification. The MEM analysis was restricted to estimate the effects of MGFePP and NaFeEDTA in 2 separate categories depending on the baseline SF concentration of subjects. The effect of iron fortification on body iron stores in women in the 3 groups by baseline SF is shown in Figure 5. In women with iron deficiency at baseline (SF concentration <15.0 μg/L), there was a significantly greater increase in body iron stores for NaFeEDTA than for MGFePP fortification ($P < 0.05$, linear effect). The difference in body iron stores between the 2 iron compounds was not significant for subjects with low iron status (SF concentrations of 15.0–24.9 μg/L) at baseline.

Comparison of efficacy estimates of random serial sampling, sparse random serial sampling, and conventional fixed time-point sampling

Efficacy estimates for the effect of MGFePP and NaFeEDTA fortification on body iron stores in women obtained by random serial sampling and simulated sparse random serial sampling at 32 wk are summarized and compared with fixed time-point sampling in Table 3. The estimates are presented as group differences (in mg Fe/kg body weight) for MGFePP and NaFeEDTA compared with the placebo and as group differences in the prevalence of elevated TfR concentrations.
between MGFePP and NaFeEDTA. Overall, the group differences of the straight-line and linear-term MEM efficacy estimates of body iron stores at 32 wk from the sparse sampling models differed only marginally from the 5-time-point random sampling estimate. The performance of sparse sampling in estimating absolute changes in body iron stores was thus comparable with the full serial sampling. Compared with the placebo, significance testing for fixed time-point sampling by using ANCOVA did not detect an increase in body iron stores for MGFePP, whereas the slope approach from the random serial sampling was more sensitive and was able to detect the small increase in body iron stores ($P < 0.05$). The efficacy estimate from sparse random serial sampling with 3 observations obtained by bootstrapping did not reach significance but indicated a trend for improved body iron stores for the MGFePP group ($P = 0.0971$). The effect of NaFeEDTA on body iron stores compared with the effect of the placebo was detected by all 3 sampling approaches [ie, random serial sampling ($P < 0.0001$), simulated sparse serial sampling ($P < 0.0001$), and fixed time-point sampling ($P < 0.001$)]. The sparse random serial estimates showed that MGFePP and NaFeEDTA fortification were efficacious in 38% and 99% of 100 bootstrap replicates, respectively. The difference in body iron responses between the MGFePP and NaFeEDTA groups at 32 wk was significant for the random serial sampling ($P < 0.05$) and the fixed time-point analysis ($P < 0.05$) but not for the sparse random serial sampling. The comparison of linear terms from the quadratic models with the simple straight-line approach showed that the linear term estimates were nearly double the size. This indicates that increases in body iron might have been faster in the early phase of the study.

FIGURE 4. Differences in measured prevalence [in percentage units (% units)] at 32 wk compared with baseline of the effects of iron fortification on the prevalence of anemia (hemoglobin concentration <12 g/L), iron deficiency (serum ferritin concentration <15 μg/L), elevated transferrin receptor (TfR; TfR concentration >4.4 mg/L), and depleted body iron stores (<0 mg Fe/kg body weight) in women in the 3 groups over 32 wk. MGFePP, micronized ground ferric pyrophosphate; NaFeEDTA, sodium iron edetate. \(^a\)\(^b\)Difference from placebo (generalized linear model for binomial data): \(^aP < 0.05\), \(^bP < 0.001\).

FIGURE 5. Effect of iron fortification on body iron in women in the 3 groups by baseline serum ferritin (SF) concentrations. Graphs were derived from the fixed part of the mixed-effects regression model for repeated measurements over 32 wk. MGFePP, micronized ground ferric pyrophosphate; NaFeEDTA, sodium iron edetate. \(^a\)Difference from MGFePP, $P < 0.05$ (linear effect).
TABLE 3 Comparison of the effect of iron fortification with micronized ground ferric pyrophosphate (MGFePP) or sodium iron edetate (NaFeEDTA) with body iron stores in women over 32 wk obtained from longitudinal random sampling (5 samples), simulated sparse random sampling (2 samples), and fixed time-point sampling (2 samples).

<table>
<thead>
<tr>
<th></th>
<th>MGFePP vs placebo</th>
<th>NaFeEDTA vs placebo</th>
<th>MGFePP vs NaFeEDTA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daily change in body iron (mg Fe/kg body weight)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random serial sampling (MEM)</td>
<td>0.0065 (0.0026, 0.0117)</td>
<td>0.0971 (0.0001, 0.7203)</td>
<td>0.0138 (0.0107, 0.0189)</td>
</tr>
<tr>
<td>Sparse random serial sampling (MEM)</td>
<td>0.0001 (0.0000, 0.0017)</td>
<td>0.0073 (0.0017, 0.0125)</td>
<td>0.0652 (0.0009, 0.8443)</td>
</tr>
<tr>
<td>Fixed time-point sampling (ANCOVA)</td>
<td>0.0001 (0.0000, 0.0017)</td>
<td>0.0065 (0.0017, 0.0125)</td>
<td>0.0436 (0.0009, 0.8443)</td>
</tr>
</tbody>
</table>

DISCUSSION

This study shows that iron fortification of margarine with MGFePP or NaFeEDTA improved body iron stores in women of reproductive age. We showed that a new efficacy study design that uses a longitudinal random sampling strategy and MEM data analysis obtained more robust efficacy estimates of the impact of iron fortification than did a conventional fixed time-point evaluation and traditional statistical methods (ANCOVA).

The random serial-sampling design has several advantages over conventional study designs in the evaluation of the efficacy of iron fortificants. First, the number of study subjects can be sharply reduced (11, 14, 15, 31). The sample size calculations in the current study were made to detect a change in SF concentrations of 2.5 μg/L. Compared with a fixed time-point study design, serial sampling reduced the required sample size from 320 to 34 subjects per group. Second, random sampling makes the timing of sample collection more flexible, and the blockwise random sampling technique allows for rescheduling of individual contacts within each block (32). This flexibility provides investigators room to space out sampling, is appealing to subjects with busy and/or unpredictable schedules, and may reduce the number of dropouts and increase compliance. The use of a sparse-sampling scheme makes studies less invasive, which is of particular importance in pediatric studies in which the number of blood samples must be minimized.

The clear advantages of MEMs in efficacy evaluation are also illustrated in this study. MEMs account for the between repeated measures and assume equal covariance of all data points and account for unequal sampling intervals and for unbalanced and/or incomplete observations (29, 33). Therefore, all obtained data points can be included in data analyses and contribute to the effect estimates even if a subject withdraws early from the study. Moreover, MEMs simultaneously determine both the within- and between-person factors that contribute to the change in iron status by examining the heterogeneity attributed to both population and individual variations (12, 32, 34). Thus, they may also have advantages in evaluating iron fortificants in larger population-based effectiveness studies where interindividual variation may be greater. A limitation of these new approaches to evaluate efficacy is that they demand skilled statisticians for data analyses, although standard statistical software can be used.

The current study showed that low-fat margarine forms a new potential vehicle for iron fortification. Margarine and low-fat spreads have previously successfully been fortified with vitamin A and vitamin D (35), vitamin E (36), folic acid (37), probiotics (38), plant sterols and stanols (39), and ω-3 fatty acids (40). The addition of iron to polyunsaturated fat can promote peroxidation and consequently result in adverse sensory changes. Compartmentalization by adding the iron to the water phase of low-fat margarine may allow iron fortification of fat-based spreads while minimizing lipid oxidation and sensory changes, as was shown in the current study with water-insoluble MGFePP or as an EDTA chelate. However, we did not study the stability of...
these products during long-term storage that exceeded the study requirements, and we did not strictly evaluate the consumer acceptability of the tested products.

To our knowledge, our study is the first iron-efficacy study performed in a European population that demonstrated a positive effect. The efficacy of MGFePP fortification in this study to increase body iron stores was 2–3 times lower than for NaFeEDTA fortification. This finding is in agreement with the estimated relative bioavailability, compared with that of ferrous sulfate, of 21–75% for MGFePP and 200–400% for NaFeEDTA from noninhibitory meals (3). MGFePP is a poorly water-soluble iron compound and must be first dissolved in the gastric juice and reduced from ferric to ferrous iron before being absorbed (41). The calculated absorption derived from the overall change in body iron stores was 2.8% for MGFePP and 5.9% for NaFeEDTA in this study, which was comparable with single-meal isotope studies that have reported absorption rates \( \approx 1.7–5.5\% \) for MGFePP (42–44) and \( \approx 3.9–16.8\% \) for NaFeEDTA (45). However, the calculated absorption generated from efficacy studies may provide a more realistic estimate of true absorption from a balanced diet (4), although this has never been directly compared. The absorption of MGFePP in the current study was in agreement with the \( \approx 1–3\% \) absorption reported from earlier efficacy studies conducted in African and Indian children (46–50). The lower absorption of MGFePP previously observed may have been due to a lower bioavailability diet and higher infection rates, which can reduce iron absorption and utilization (51).

This is the first study to evaluate the efficacy of MGFePP fortification (mean particle size: \( \approx 2.5\mu m \)) in women. The observed improvement in body iron stores was not large enough to significantly reduce the prevalence of iron deficiency in this population with low anemia prevalence. A study with larger sample size or longer study duration may detect an effect on prevalence. Hotz et al (52) tested a microencapsulated form of ferric pyrophosphate (FePP) with a smaller particle size (mean particle size: 0.3 \( \mu m \)) in extruded rice in Mexican women with a high prevalence of anemia over 6 mo. Similar amounts of fortification iron were consumed as in our study, and the changes in body iron stores were comparable between the 2 studies. The calculated absorption of FePP in the Mexican study (3%) was in the same range as in our study. The prevalence of iron deficiency was significantly reduced in the Mexican women with high anemia prevalence. The estimation of efficacy of NaFeEDTA in our study is consistent with an earlier efficacy study of fortified fish sauce in women (53) and effectiveness studies of curry powder, fish sauce, and sugar (54–56).

Iron absorption is inversely related to iron status, and bioavailability of iron from water-soluble compounds, such as ferrous sulfate and NaFeEDTA, is up-regulated in women with low iron stores (43, 57–59). Earlier intervention studies of NaFeEDTA (54) reported the largest increment in body iron stores in the most iron-deficient women. In contrast, in isotope-labeled test meals, there was only minimal up-regulation of absorption of poorly water-soluble iron compounds, such as FePP, in subjects with poor iron status (43). This differential up-regulation of iron absorption from different iron compounds has not been previously compared in an efficacy study. Our data clearly confirmed the effect: in women with iron deficiency at baseline, there was a significantly greater increase in body iron stores for the NaFeEDTA group than for the MGFePP group, which was a difference not shown in iron-sufficient subjects (Figure 5).

In conclusion, low-fat margarine fortified with MGFePP and NaFeEDTA improved the iron status in young women with low iron stores. Random serial sampling is a promising method to increase sensitivity, simplify logistics, and reduce costs of intervention studies that evaluated the effect of iron fortification.

We thank all of the women in the study for their participation. The authors’ responsibilities were as follows—MA, WT, MBZ, MJ, GSMJED, LGJF, and RFH: designed the study; MA and JTF: conducted the study and performed laboratory analyses; MA and WT: conducted the statistical analyses; MA: wrote the first draft of the manuscript; and all authors: contributed to data interpretation and edited the manuscript. MA, JTF, MBZ, and RFH declared no conflicts of interest. WT, MJ, GSMJED, and LGJF are employed at Unilever Research and Development.

REFERENCES

37. Mathioudakis B. Recommendations for vitamin intake in the European... 75:6198–204.


35. Roy A, Ette EI. A pragmatic approach to the design of population... 1104 ANDERSSON ET AL


