Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Ross JL, Quigley CA, Cao D, et al. Growth hormone plus childhood low-dose estrogen in Turner's syndrome. N Engl J Med 2011;364:1230-42.

GH and Childhood Low-dose Estradiol in Turner Syndrome:

A Placebo-Controlled Trial

-- Supplementary Appendix --

Judith L. Ross, MD[1]; Charmian A. Quigley, MBBS[2]; Dachuang Cao, PhD[2];
Penelope Feuillan, MD[3†]; Karen Kowal, PA[1]; John J Chipman, MD[4];
Gordon B. Cutler Jr, MD[4,5,6]

[1]Thomas Jefferson University, Philadelphia, Pennsylvania, 19107 and A.I. DuPont Hospital for Children, Wilmington, Delaware, 19803.

[2]Lilly Research Laboratories, Indianapolis, Indiana, 46285.

[3]National Institute of Child Health, and Human Development, Bethesda, Maryland, 20892,[†]Deceased

[4]Lilly Research Laboratories, Indianapolis, Indiana, 46285, retired.

[5]National Institute of Child Health, and Human Development, Bethesda, Maryland, 20892, retired

[6]Current address: 630 Bland Point Rd, Deltaville, Virginia 23043

Corresponding Author:

Judith L. Ross, MD
Thomas Jefferson University
Jefferson Medical College Department of Pediatrics
1025 Walnut St.
Philadelphia, PA 19107 USA
Phone: +1 215 955 1648
Email: judith.ross@jefferson.edu

Supplementary Appendix: Table of Contents

Abbreviations: EE2 = ethinyl estradiol; GH=growth hormone; LDE = low-dose estrogen (during childhood); OE = oral estrogen; OP = oral placebo.

1. Method for preparation of ethinyl estradiol solution

This information is reproduced from the method provided by the Pharmaceutical Development Branch of the Clinical Center at the National Institutes of Health

Drug: Ethinyl estradiol solution 1 mcg/mL

Batch Size: 40,000 mL (83 x 480 mL)

Storage Condition: Controlled room temperature (15°C to 30°C)

Ingredients and Quantities Required for this 2-step Process

Ingredient	Quantity Per Unit	Required Batch Quantity
Step 1 (dissolve ethinyl estradiol in ethanol)		
Ethinyl estradiol	1 mcg*	42 mg
Ethanol (190°)	0.001 mL	40 mL
Step 2 (dissolve benzoic acid in ethanol/water solution)		
Benzoic acid, USP	1 mg	40 gm
Ethanol (190°)	0.005 mL	200 mL
Water for injection	1 mL	40 L

*Includes 5% overage to account for filtration loss. USP = United States Pharmacopeia

Procedures

Step 1	Dissolve ethinyl estradiol in ethanol
1.	Weigh the required amount of ethinyl estradiol (42 mg for 40 L batch)
2.	To a clean 50 mL beaker add a magnetic stirrer
3.	Add ethinyl estradiol to beaker
4.	Add 30 mL ethanol to beaker and mix until dissolved. ***Mix time:*** ~10 minutes.
Step 2	**Dissolve benzoic acid in ethanol/water solution**
1.	Calibrate a 40 L carboy
2.	Add 36 L water for injection
3.	Add 200 mL ethanol to the carboy; mix until clear using a propeller blade. ***Mix time:*** ~30 minutes.
4.	Weigh the benzoic acid, and record the weight.
5.	Draw up the dissolved ethinyl estradiol in a syringe and filter directly into the carboy using a 5-micron Monoject syringe. Add an extra10 mL ethanol, and mix to remove any additional drug. Filter and add to the carboy.
6.	Add water for injection to 39 L and mix. ***Mix time:*** ~30 minutes.
7.	Measure the pH. Adjust if necessary to pH 3.13 (range 2 – 4).
8.	Add water for injection as needed to final volume 40 L and ***mix*** for an additional 30 minutes.
9.	Using the Wheaton pump, adapt a Silastic tube, and calibrate it to deliver 240 mL at a time.
10.	Aliquot 480 mL into each amber plastic bottle and store at room temperature.

2. Reasons for discontinuation from study prior to protocol completion

Reasons for discontinuation prior to protocol completion and non-inclusion in the adult height (AH) population (n=58) in order of decreasing frequency were: patient decision, n=34 (e.g., to pursue open-label GH treatment; issues with injections; travel problems; advanced age or satisfaction with height [16/34 were >14 years old at discontinuation]); transferred to open-label GH after data and safety monitoring board closed study, n=7); protocol violation, n=7; adverse events, n=4; loss to follow-up, n=4; physician decision, n=2 (1 because of newly diagnosed Sanfilippo syndrome, 1 because of non-compliance).

3. Treatment compliance

Compliance with study drug injections (defined as receipt of ≥80% of expected injections) was 83% overall and was similar among treatment groups. Oral study drug compliance relative to prescribed dosages was 98% overall and was similar among treatment groups.

4. Data for patients randomized to placebo injection groups who later received GH treatment

a. Methods

As a prospectively defined condition of analysis, patients in the placebo injection groups who discontinued the study before protocol completion and indicated at post-study follow-up that they had subsequently received open-label GH treatment were excluded from the efficacy analyses, but retained in the safety analyses. Data from these 8 patients are summarized below.

One additional placebo-treated patient (P/E group) was discovered to have received surreptitious GH treatment for 15 months while participating in the study, but her treatment status was unknown at the time of the analyses. Therefore, her data are included in the analyses of the group to which she was randomized.

b. Results

When the analyses of the intent-to-treat (ITT) population were re-run with inclusion of the 8 patients who received GH after leaving the study there was no change in the estimate of efficacy (**Supplementary Table 1**). Similarly, when the analyses of the ITT population were re-run without the patient who received surreptitious GH during the study, there was no change in the estimate of efficacy.

Supplementary Table 1: Patients who received placebo injections during study and GH treatment after study, and returned for post-study follow-up

	Baseline		Last On-Study		Post-Study Follow-Up	
	P/P (n=5)	P/E (n=3)	P/P (n=5)	P/E (n=3)	P/P (n=5)	P/E (n=3)
Chronological age (y)	7.2±1.9	7.3±3.4	9.9±2.1	12.6±1.2	16.5±4.6	16.2±3.6
Bone age (y)	6.2±2.1	4.6±0.6 (n=2)	8.4±2.3	12.5±0.7 (n=2)	15.2±3.8	12.7±0.6
Time on study (y)	NA	NA	2.7±1.9	5.3±4.6	9.2±3.9	8.9±0.3
Height (cm)	104.8±8.8	111.4±22.3	116.3±7.6	135.3±4.8	140.4±13.1	141.5±6.7
Height SDS	-3.4±0.5	-2.0±1.3	-3.2±0.3	-2.7±1.4	-2.6±0.5	-3.1±0.8

Abbreviations: P/P=placebo injection/oral placebo; P/E=placebo injection/childhood oral low dose estrogen (LDE); NA = not applicable; SDS = standard deviation score; y = year

5. Oral study drug dosage individualization

a. Methods

i. Study procedures

The protocol specified that the oral study drug dosage (ethinyl estradiol [EE2] or placebo [P]) could be reduced by 50% at the discretion of the investigator if any of the following occurred: breast development ≥Tanner stage 2 before chronological age 12 (defined as premature breast development); vaginal bleeding before age 14 (defined as premature vaginal bleeding); bone age advancement of 2 years per chronological year or bone age greater than chronological age up to age 14. If the dosage was reduced, then at the next protocol-specified dosage increase the patient's current dose was doubled, but the dose remained below the protocol-specified dosage for age thereafter.

ii. Statistical analyses

To determine whether a dosage reduction had occurred, the patient's prescribed oral study drug dosage in ng/kg/day at a given visit was compared with the protocol-specified dosage for that age. If the patient's prescribed dosage was ≤ 60% of protocol-specified dosage, that visit was flagged as a dosage reduction. Each patient was flagged as "reduced" at the first visit at which a reduction occurred, so as to count each patient only once. To calculate the percentage of patients who underwent dosage reduction at a particular dosage level, the number of patients who had a first reduction at that dosage level was divided by the number of patients within that dosage level who had not had a dosage reduction at a previous dosage level (i.e., those who initiated the protocol-specified dosage for age). The number of patients who underwent reductions during the childhood phase of the study (25 and 50 ng/kg/day dosage levels) was calculated as the total number of patients who had an initial dosage reduction at either of these 2 dosage levels.

Based on review of reported events, the reasons for oral study drug dosage reduction were summarized within 5 categories: premature breast development; premature vaginal bleeding; bone age advance; changes in emotion, mood or behavior; headache, and other reasons. The number and percentage of patients

whose dosage was reduced for a given reason were summarized by treatment group and also by the pooled groups of patients who were randomized to receive oral placebo during childhood (P/P + GH/P = OP) *vs.* oral LDE during childhood (P/E + GH/E = OE). In addition, the reasons for initial reduction were summarized at each dosage level for the number and percentage of patients by pooled treatment groups, who had not had a reduction at a prior dosage level, again, to count each patient only once.

Average EE2 or oral placebo dosages by year-on-study were calculated by treatment group (**Supplementary Figure 1**). However, because ages at study entry ranged from 5 to 12.9 years, each annual timepoint could encompass as much as a 7.9-year age span (until patients started to complete the study at maturity), and therefore a wide range of potential protocol-specified (and actual) oral study drug dosages. For example, in year 2 of the study the protocol-specified EE2 dosage could have ranged from 25 to ≥ 200 ng/kg/d; year 3 from 25 to ≥ 400; year 4 from 25 to ≥ 800. However, because of dose reductions as a result of the dosage individualization protocol, the mean age-related dose did not exceed 150 ng/kg/d up to age 16. These complexities should be considered in interpreting the average oral drug dosages by year-on-study as they do not reflect the exposures of individual patients.

b. Results

Because of the oral drug dose-individualization regimen, 95/149 (64%) girls (oral placebo [OP], 43/74 [58%]; oral estrogen [OE], 52/75 [69%], p=0.18) underwent one or more reductions of their protocol-specified EE2 (or placebo equivalent) dosages over the course of the study, during both the prepubertal and pubertal phases combined. Fifty of the 95 girls underwent more than one dosage reduction (50/149 [34%] total; OP, 28%; OE, 39%; p=0.23).

Supplementary Tables 2a and 2b present summaries of the reasons for reduction of oral study drug by randomized treatment group and by pooled treatment group (OP *vs.* OE) at the protocol-specified dosage levels.

Reductions at the childhood dosage levels were: 25 ng/kg/day, OP, 11% *vs.* OE, 38% p=0.011; 50 ng/kg/day, OP, 39% *vs.* OE, 54%, p=0.17, primarily because of premature breast development (OP [11%] *vs.* OE [44%], p<0.001). Overall, 40% of girls receiving OP vs. 59% of girls receiving OE (p=0.036) underwent initial dosage reductions during the childhood phase of the study.

The finding of spontaneous pubertal development (i.e., in the absence of estrogen supplementation) has a number of potential explanations. While it might have been expected that spontaneous breast development would have been more commonly associated with a mosaic karyotype, in fact only 2 of the 9 girls who had early breast development without estrogen had karyotypes other than 45,X monosomy. Possible alternative reasons for the finding of apparent spontaneous pubertal development in this study cohort include:

(1) *Observer bias:* girls with TS are rarely examined so closely and longitudinally for breast development, so perhaps there was truly a higher rate of initial breast development than previously reported, as a result of residual ovarian estrogen secretion combined with aromatization of adrenal

androgens. Furthermore, because we were intent on minimizing early feminization of our patients, we may have overcalled some cases of early thelarche, particularly in overweight patients in whom the distinction between subcutaneous fat and true breast tissue may be difficult.

(2) *Secular trend:* during the time the study was being performed, there was a secular trend toward earlier thelarche, accompanied by a rising prevalence of childhood obesity, which can accelerate GnRH activation and adrenarche. These factors might underlie a true increase in spontaneous breast development in our patients, as earlier central activation may have occurred before completion of X-monosomy-induced ovarian follicle depletion.

During the pubertal induction phase after age 12, when all patients were receiving escalating EE2 replacement (protocol dosages increasing from 100 ng/kg/day for all groups), initial reductions were made for 30/65 (46%) girls who had not had prior reductions (i.e., those who initiated the protocol-specified dosage for age), most commonly for premature vaginal bleeding (before 14 years) and bone age advancement at the 100 ng/kg/day dosage. Because many girls were already on reduced EE2 dosages by the time they reached ages at which they would have received the protocol-specified pubertal dosages of 200 and 400 ng/kg/day, initial dosage reductions were uncommon at these levels. However, the actual dosages received at the 200 and 400 ng/kg/day protocol levels were only approximately 40-50% of protocol-specified dosages (as shown in Figure 1b, main manuscript).

Median age at Tanner stage 2 breast development was 12.1 years for the girls who received childhood LDE (combined P/E and GH/E groups) versus 12.8 years for girls who received childhood oral placebo (combined P/P and GH/P groups; p=0.062).

Mean ages at menarche were 14.46±1.09 and 14.93±1.14 years at mean EE2 dosages of 87±54 ng/kg/day and 107±82 and ng/kg/day for the pooled OP and OE groups, respectively (p=0.17).

No patient received the protocol-specified EE2 dosage of 800 ng/kg/day (~40 mcg/day); however, oral contraceptives containing 30 mcg EE2 were initiated in 64% of patients at an average age of 16.3±0.8 years. Of note, the lowest commercially available dosage of ethinyl estradiol is 20 mcg, which is higher than the dosages used for most patients in this study.

Although 1 patient in the P/E group reported dysfunctional uterine bleeding, no patient discontinued the study on the basis of premature puberty or other estrogen-related effect.

Supplementary Table 2: Summary of reasons for oral study drug dosage reductions for all randomized patients

a. Number (percent) of patients with dosage reductions by treatment group

Reason for Dosage Reduction	P/P (n=39)	P/E (n=40)	GH/P (n=35)	GH/E (n=35)	OP (n=74)	OE (n=75)	P-value* OP vs. OE
Premature breast development	6 (15%)	17 (43%)	6 (17%)	9 (26%)	12 (16%)	26 (35%)	0.014
Premature vaginal bleeding	0 (0%)	3 (8%)	2 (6%)	5 (14%)	2 (3%)	8 (11%)	0.098
Bone age advance	9 (23%)	3 (8%)	11 (31%)	4 (11%)	20 (27%)	7 (9%)	0.006
Headache	3 (8%)	4 (10%)	0 (0%)	2 (6%)	3 (4%)	6 (8%)	0.494
Emotion/mood/behavior	1 (3%)	2 (5%)	0 (0%)	1 (3%)	1 (1%)	3 (4%)	0.620
Other	3 (8%)	0 (0%)	2 (6%)	2 (6%)	5 (7%)	2 (3%)	0.276
Total patients dose reduced	**22 (56%)**	**29 (73%)**	**21 (60%)**	**23 (66%)**	**43 (58%)**	**52 (69%)**	**0.175**

P/P = Placebo/Placebo; P/E = Placebo/Estrogen; GH/P = GH/Placebo; GH/E = GH/Estrogen; OP = Oral Placebo (P/P + GH/P); OE = Oral Estrogen (P/E + GH/E). Only first event leading to dosage reduction is reported, so that each patient is counted only once. Within-group percentages are shown. Patients who underwent a dosage reduction at any time typically continued to receive lower than protocol-specified dosages thereafter. *P-value from Fisher Exact Test

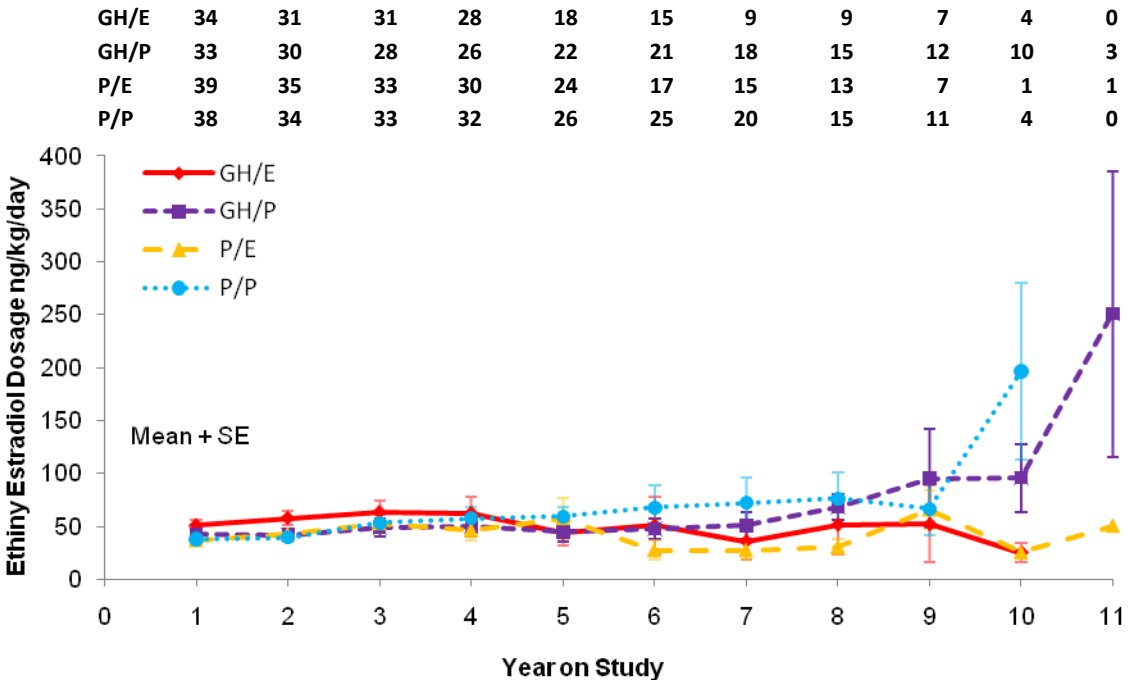
b. Number (percent) of patients by pooled treatment group by protocol-specified EE2 dosage level

Protocol Dosage Level (ng/kg/day)	25 (OP) N=44	25 (OE) N=32	P-value[a]	50 (OP) N=56*	50 (OE) N=48*	P-value[a]	100 (OP) N=37*	100 (OE) N=28*	P-value[a]	200 (OP) N=17*	200 (OE) N=19*	P-value[a]	400 (OE) N=14**
Age Range (y)		**5-8**			**>8-12**			**>12-14**			**>14-15**		**>15**
Reason for Reduction: n (%)													
Premature breast development	3 (7)	5 (16)	0.270	6 (11)	21 (44)	<0.001	3 (8)	0	0.253	0	0	1.000	0
Premature vaginal bleeding	0	1 (3)	0.421	0	0	1.000	2 (5)	5 (18)	0.224	0	2 (11)	0.487	0
Bone age advance	2 (5)	2 (6)	1.000	11 (20)	4 (8)	0.161	7 (19)	0	0.016	0	1 (5)	1.000	0
Headache	0	1 (3)	0.421	2 (4)	1 (2)	1.000	0	2 (7)	0.182	1 (6)	1 (5)	1.000	1 (7)
Emotion/mood/behavior	0	3 (9)	0.071	1 (2)	0	1.000	0	0	1.000	0	0	1.000	0
Other	0	0	1.000	2 (4)	0	0.498	2 (5)	0	0.502	1 (6)	1 (5)	1.000	1 (7)
Total	**5 (11)**	**12 (38)**	**0.011**	**22 (39)**	**26 (54)**	**0.168**	**14 (38)**	**7 (25)**	**0.299**	**2 (12)**	**5 (26)**	**0.408**	**2** (14)

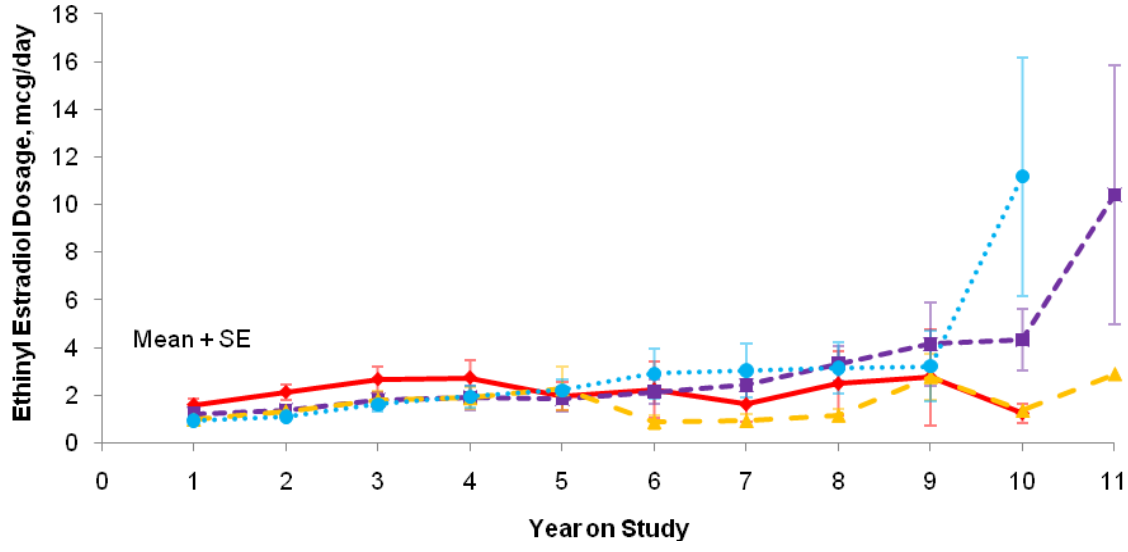
OP = Oral Placebo (P/P + GH/P); OE = Oral Estrogen (P/E + GH/E). Only first event leading to dosage reduction is reported, so that each patient is counted only once. *N=number of patients in group who had at least one visit at this dosage level and had not had a prior dosage reduction at a lower dosage level. Within-group percentages are shown. **No data are shown for OP at the 400 ng/kg/day dosage level because no patient originally randomized to the GH/P or P/P groups had a first reduction at the 400 ng/kg/day dosage level. P-values are from Fisher Exact Tests.
[a]P-values are for OP vs. OE at each EE2 dosage level

As shown in **Supplementary Figures 1a and 1b**, mean EE2 dosages by year-on-study were below 100 ng/kg/day, and 4 µg/day, respectively, for the first 8 years of study in all groups. Notably, each annual timepoint includes girls whose ages differ by up to 7.9 years and mean EE2 (or placebo, during the childhood phase) dosages vary according to this factor and the frequency of dose reductions. Numbers above the figure represent patient numbers by treatment group at each annual timepoint.

Supplementary Figure 1a: Ethinyl estradiol dosage by year-on-study, ng/kg/day



Supplementary Figure 1b: Ethinyl estradiol dosage, mcg/day



4. Safety and laboratory evaluations

a. Methods

i. Study procedures

Safety was evaluated at each study visit by adverse event history, physical examination, and laboratory analyses. A fasting blood sample was drawn at 6-month intervals (2–3 d after study drug injection) for chemistry; liver function tests; hematology; insulin-like growth factor-I (IGF-I); thyroid function tests (thyroid stimulating hormone [TSH], thyroxine [T4], free T4 and triiodothyronine [T3]; lipids (total cholesterol, HDL cholesterol, LDL cholesterol, VLDL cholesterol, triglycerides); and glucose metabolism parameters (glucose, insulin, glycosylated hemoglobin).

ii. Statistical analyses

Because the study was powered for efficacy rather than safety endpoints, no significance testing was planned *a priori*. However, post-hoc analyses using Fisher exact tests were performed comparing treatment-emergent adverse events (defined as any event that began or worsened after study entry) for the pooled GH-treated vs. placebo injection groups (GH/P + GH/E vs. P/P + P/E), and for the pooled childhood oral low-dose estrogen vs. oral placebo groups (P/E + GH/E vs. P/P + GH/P). False discovery rate adjusted P-values were calculated to control the overall type 1 error for multiple hypothesis testing. Mean values for laboratory parameters were compared among treatment groups by ANOVA at each study visit and for change from baseline to endpoint. The Quantitative Insulin Sensitivity Check Index (QUICKI) was calculated as 1/[log(fasting insulin μU/mL) + log(fasting glucose mg/dL).

b. Results

i. Serious adverse events

No deaths occurred during the study. Serious adverse events were defined by regulatory criteria as any event that: resulted in death, was life-threatening, required hospitalization or prolongation of existing hospitalization, resulted in permanent disability, resulted in a congenital anomaly in an offspring of a study patient, or was otherwise serious in the judgment of the investigator. Serious adverse events were reported for 27 of the 149 patients (18%), as shown in **Supplementary Table 3**. Only one of these events was considered likely to have been related to stud drug exposure: dysfunctional uterine bleeding in one patient in the P/E group.

ii. Treatment-emergent adverse events

Treatment-emergent adverse events (TEAEs) were reported for all 149 randomized patients (**Supplementary Table 4**). The most common event categories (listed in descending frequency overall, as range of percent for the 4 treatment groups) were: respiratory and upper airway disorders (89-94%), gastrointestinal disorders (77-86%); ear disorders (66-82%) and headache (65-82%). Of note, in contrast to previous studies (references 9 and 12, main manuscript), the prevalence of otitis media was equivalent in the GH (43%, 43%) and placebo injection groups (44%, 45%).

Post-hoc analyses were performed by Fisher exact tests comparing rates of TEAE categories for pooled placebo injection (PI) vs. GH-treated (GH) groups, and for pooled oral placebo (OP) vs. childhood oral estrogen (OE) groups. After correction for false discovery rate there were no TEAE categories that were significantly more frequent for GH-treated than placebo-injection groups or for OE than OP groups (**Supplementary Table 4**). Notably, because the study was not powered for

Supplementary Table 3: Serious Adverse Events (SAE), All Randomized Patients

P/P (n=39)	P/E (n=40)	GH/P (n=35)	GH/E (n=35)
Patients with SAE = 9	**Patients with SAE = 7**	**Patients with SAE = 4**	**Patients with SAE = 7**
1. Dehydration, fever and tonsillitis 2. Dehydration due to gastroenteritis	1. Gastritis 2. Dysfunctional uterine bleeding	1. Cellulitis right foot 2. Balloon angioplasty for mitral stenosis	Bilateral salpingo-oophorectomy (Y-chromosomal fragment detected in peripheral karyotype after study entry)
Cardiac catheterization and repair of coarctation of the aorta	Nasal polypectomy	Aortic valve replacement	1. Multiple operations to repair 2 events of traumatic compound fractures of both arms. 2. Debridement of suture abscess
Dehydration due to gastroenteritis	Needle aspiration drainage of septic knee joint	Dehydration due to diarrhea and vomiting resulting from Munchausen syndrome by proxy (mother had been giving laxatives)*	Intravenous antibiotic treatment of otitis media
Tympanoplasty	Repair of coarctation of the aorta	1. Cardiac catheterization and attempted balloon angioplasty for aortic stenosis 2. Aortic valve replacement 3. Tympanoplasty 4. Transfusion for postoperative bleeding	1. Thrombocytopenia due to pre-existing Sanfilippo syndrome diagnosed after study entry[a] 2. Pneumonia and anemia
1. Tonsillectomy and adenoidectomy 2. Unilateral leg edema[a] 3. Cardiac catheterization and balloon dilatation for aortic stenosis 4. Worsening aortic stenosis, aortic valve replacement, closure of patent foramen ovale, pacemaker placement*	1. Dehydration due to vomiting and diarrhea* 2. Abdominal pain and dehydration due to vomiting and diarrhea		1. Cardiac catheterization 2. Aortic valve replacement
1. Multiple surgeries for cholesteatoma of both ears 2. Mastoidectomy 3. Tympanoplasty	1. Elective orthodontic surgery 2. Pyelonephritis 3. Dehydration due to influenza, vomiting and urinary tract infection		1. Cardiac catheterization 2. Repair of coarctation of the aorta
Pneumonia	Posterior spinal fusion with rod placement for scoliosis, judged as possibly related to treatment		Cellulitis of the leg
Osteotomy of mandible and removal of molars			
1. Von Willebrand disease[a] 2. Post-tonsillectomy hematoma			

Serious adverse events were defined according to regulatory criteria as any event that: resulted in death, was life-threatening, required hospitalization or prolongation of existing hospitalization, resulted in permanent disability, resulted in a congenital anomaly in an offspring of a study patient, or was otherwise serious in the judgment of the investigator. Each cell in the table represents a single patient; a number of patients had multiple SAEs. Except as noted, all events were designated as serious on the basis of hospitalization. [a]Serious according to investigator for other reason; *Resulted in discontinuation from study. Only the event of dysfunctional uterine bleeding in a patient in the P/E group is considered likely to have been related to study drug exposure.

Supplementary Table 4: Treatment-emergent Adverse Events, All Randomized Patients

Numbers shown as percent of group	P/P (n=39)	P/E (n=40)	GH/P (n=35)	GH/E (n=35)	Total (n=149)	P-value PI vs. GH[1]	P-value OP vs. OE[1]
Any Event	**100**	**100**	**100**	**100**	**100**	**1.000**	**1.000**
Adverse drug reaction*	12.8	7.5	5.7	5.7	8.1	0.379	0.563
Allergy	10.3	22.5	17.1	31.4	20.1	0.306	0.065
Asthma	5.1	2.5	2.9	11.4	5.4	0.475	0.719
Behavioral/emotional disorders	30.8	37.5	34.3	25.7	32.2	0.603	1.000
Blood glucose elevated	0.0	0.0	0.0	2.9	0.7	0.470	1.000
Bone disorders	15.4	10.0	2.9	5.7	8.7	0.086	0.780
Breast disorders**	12.8	12.5	11.4	11.4	12.1	1.000	1.000
Cardiac disorders	25.6	25.0	20.0	11.4	20.8	0.163	0.550
Cardiac procedures	5.1	2.5	8.6	5.7	5.4	0.475	0.494
Central nervous system disorders	41.0	27.5	25.7	25.7	30.2	0.288	0.376
Ear disorders	82.1	77.5	74.3	65.7	75.2	0.188	0.449
Otitis media (subset of Ear disorders)	*43.6*	*45.0*	*42.9*	*42.9*	*43.6*	*0.870*	*1.000*
Edema	33.2	22.5	20.0	28.6	26.2	0.710	0.854
Ear, nose, throat surgery	43.6	32.5	28.6	11.4	29.5	0.020	0.074
Eye and visual disorders	53.8	45.0	45.7	40.0	46.3	0.511	0.413
Fever	64.1	70.0	74.3	60.0	67.1	1.000	0.728
Fracture	23.1	17.5	17.1	22.9	20.1	1.000	1.000
Gastrointestinal disorders	84.6	82.5	85.7	77.1	82.6	0.830	0.518
Gynecological disorders***	43.6	60.0	54.3	60.0	54.4	0.621	0.190
Headache	82.1	65.0	74.3	77.1	74.5	0.851	0.348
Hearing disturbance	15.4	17.5	8.6	5.7	12.1	0.129	1.000
Hematological disorders	15.4	15.0	2.9	14.3	12.1	0.314	0.452
Infectious disorders	48.7	40.0	40.0	31.4	40.3	0.318	0.319
Injection site problem	15.4	20.0	17.1	14.3	16.8	0.828	1.000
Injury	51.3	42.5	40.0	48.6	45.6	0.869	1.000
Joint disorders (not injury-related)	20.5	35.0	34.3	34.3	30.9	0.478	0.313
Laboratory test abnormalities	12.8	10.0	14.3	8.6	11.4	1.000	0.452
Lipid abnormalities	33.3	25.0	14.3	11.4	21.5	0.017	0.431
Liver abnormalities	12.8	15.0	14.3	14.3	14.1	1.000	1.000
Lymphoid enlargement	15.4	12.5	17.1	5.7	12.8	0.806	0.229
Nail anomalies	2.6	5.0	5.7	2.9	4.0	1.000	1.000
Neoplasia****	5.1	12.5	8.6	2.9	7.4	0.542	1.000
Nevi	25.6	17.5	14.3	28.6	21.5	1.000	0.842
Oral/dental disorders	46.2	47.5	28.6	34.3	39.6	0.066	0.663
Orthopedic disorders (other than joint)	10.3	10.0	8.6	11.4	10.1	1.000	1.000
Pain (various, no precipitant reported)	33.3	52.5	37.1	54.3	44.3	0.869	0.032
Pain due to precipitating event/condition	12.8	25.0	28.6	31.4	24.2	0.129	0.339
Renal and urinary tract disorders	23.1	40.0	28.6	34.3	31.5	1.000	0.159
Respiratory and upper airway disorders	89.7	92.5	88.6	94.3	91.3	1.000	0.401
Scoliosis	30.8	40.0	48.6	40.0	39.6	0.271	1.000
Slipped growth plate	0.0	0.0	0.0	2.9	0.7	0.470	1.000
Skin disorders	71.8	60.0	71.4	77.1	69.8	0.288	0.722
Surgical procedures various	25.6	17.5	8.6	34.3	21.5	1.000	0.319
Thyroid disorders	35.9	35.0	37.1	34.3	35.6	1.000	0.865
Toe disorders	12.8	12.5	14.3	14.3	13.4	0.813	1.000
Weight disorders	23.1	17.5	34.3	25.7	24.8	0.188	0.348

P/P = placebo injection/oral placebo; P/E = placebo injection/oral estrogen; GH/P = GH injection/oral placebo; GH/E = GH injection/oral estrogen. [1]Significance testing by Fisher exact test for pooled groups (PI=P/P+P/E; GH=GH/P+GH/E; OP=P/P+GH/P; OE=P/E+GH/E); all P-values were non-significant when adjusted for false discovery rate (multiple hypothesis testing). *Reactions to concomitant medications such as antibiotics. **Includes terms such as: breast discomfort or tenderness, breast asymmetry, breast growth, nipple enlargement or discomfort;; ***Includes terms such as: menstrual problems (irregular, intermittent, heavy etc); vaginal discharge; itching, discomfort or redness in the genital area; ****Includes lipoma, mucocele, pilonidal swelling, “skin benign neoplasm”, “tumor recurrence” (see Supplementary Appendix text section 4.b.ii) and unspecified growths, lumps, or nodules. Numbers shown are percentage of each randomized group. Events are presented in alphabetical order.

safety *a priori*, the between-group comparisons should not be interpreted as providing conclusive evidence of the safety of GH or childhood EE2.

To evaluate whether any events for individual patients were likely to have been related to study drug exposure we reviewed all TEAEs to identify any events that abated upon discontinuation of study drug and reappeared after reintroduction of study drug. One patient in the GH/E group had an event of "continuous migrating joint pain" that was reported to have resolved upon discontinuation of study drugs (oral and injectable) and reappeared after their reintroduction. However, this patient remained on-study for over 10 years, discontinuing at age 16.9 when her bone age x-ray demonstrated fused epiphyses.

Two other patients had TEAEs worthy of comment, as they had events coded as "neoplasia" or "tumor recurrence", although neither was reported as serious. One girl in the GH/P group had a benign juvenile melanoma (Spitz cell nevus) partially removed prior to study entry; when this lesion remained 1 year after study entry (reported as "tumor recurrence"), she underwent additional surgery to complete the removal. She remained in the study a further 8 years with no additional episodes of this problem reported. One girl in the GH/E group underwent skin biopsy after 2 years on study for removal of a "skin benign neoplasm" (nevus) with "atypical spindle cell proliferation". This lesion was subsequently removed by surgical excision and did not recur during the remaining 4 years of the patient's study participation.

Overall, there were no new or unexpected safety signals with respect to GH or ethinyl estradiol treatment in this study.

ii. Weight and body mass index

There were modest increases in weight SDS and body mass index from baseline to endpoint in all treatment groups, with no significant differences among groups (Table 1, main manuscript).

iii. Laboratory and radiology data

Insulin-like growth factor-I

Mean IGF-I concentrations were within the normal range at baseline. Expressed as SDS these were (by treatment group): P/P, -0.4±2.3; P/E, -0.6±1.9; GH/P, 0.1±2.0; GH/E, -0.1±2.1; p=0.52. As shown in **Supplementary Figure 2**, mean (± SE) post-baseline IGF-I concentrations were consistently greater in the two GH-treated groups than the two placebo injection groups ($p<0.001$), with similar patterns for the GH/P and GH/E groups, indicating that the addition of LDE to GH in childhood had little effect on systemic IGF-I concentrations. Interpretation of apparent differences among treatment groups toward the end of the study is hampered by small patient numbers.

Glucose metabolism parameters

Mean (± SE) values for fasting blood glucose and glycosylated hemoglobin were normal throughout the study and showed small changes from baseline to endpoint, with no significant differences among treatment groups. Insulin sensitivity assessed by fasting insulin concentrations and QUICKI declined minimally across the duration of the study, with no significant differences among treatment groups for changes from baseline to endpoint. A summary of baseline-to-endpoint changes in glucose metabolism parameters is provided in **Supplementary Table 5**.

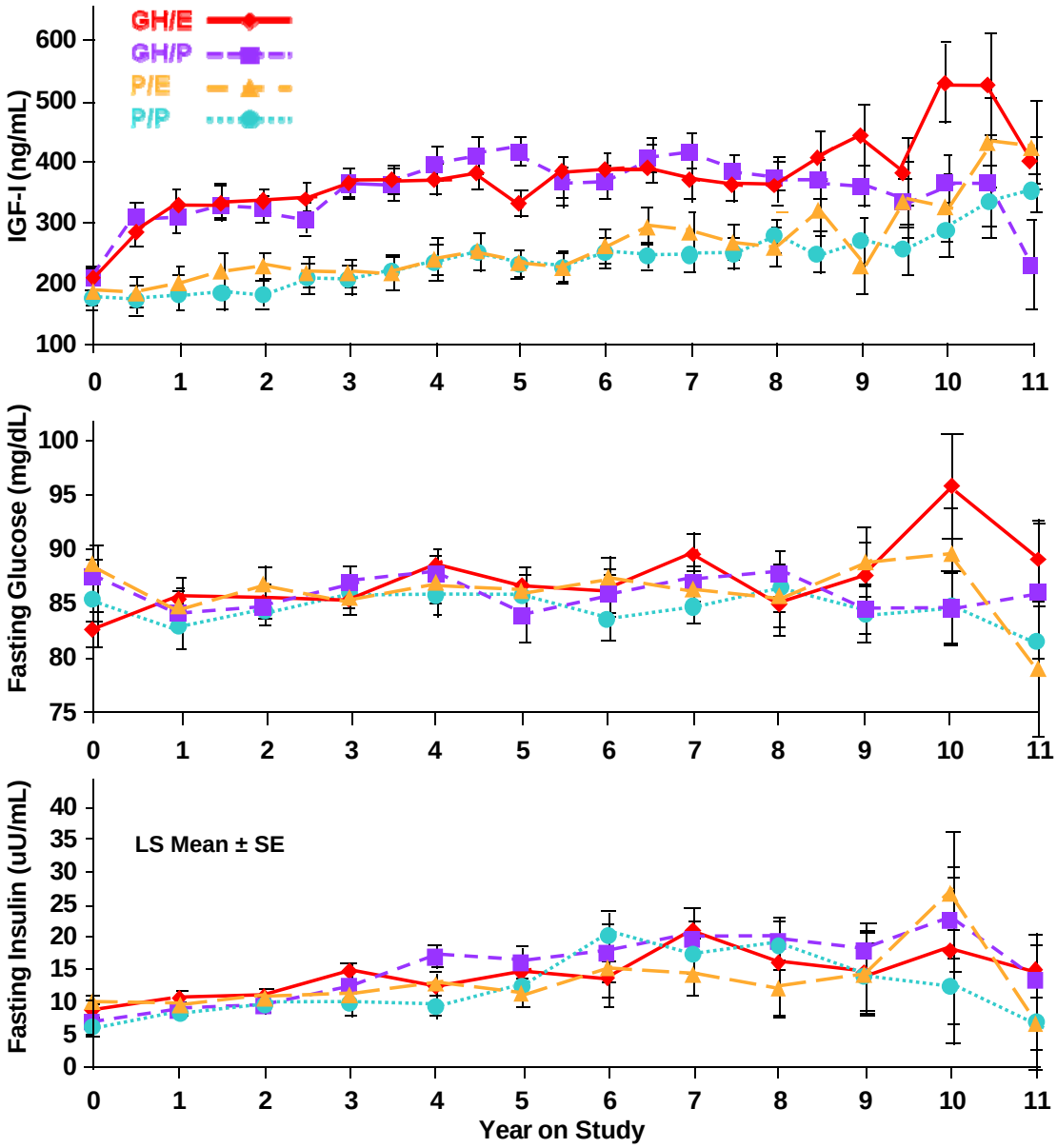
Supplementary Table 5: Summary of changes from baseline to endpoint for glucose metabolism parameters

Group	P/P	P/E	GH/P	GH/E	P-value
Fasting glucose (mg/dL)	-1.1±10.1 (24)	+1.7±12.0 (31)	-2.0±7.0 (26)	+2.7±7.8 (27)	0.233
Fasting insulin (μU/mL)	+9.0±16.7 (21)	+7.1±15.4 (28)	+6.3±8.3 (25)	+3.9±8.1 (25)	0.596
QUICKI	-0.03±0.09 (21)	-0.03±0.06 (28)	-0.03±0.08 (26)	-0.02±0.06 (26)	0.979
Glycosylated hemoglobin (%)	-0.20±0.87 (31)	+0.05±0.93 (36)	-0.19±1.16 (32)	-0.15±0.70 (32)	0.632

P/P = Placebo/Placebo; P/E = Placebo/Estrogen; GH/P = GH/Placebo; GH/E = GH/Estrogen; OP = Oral Placebo (P/P + GH/P); OE = Oral Estrogen (P/E + GH/E). Values shown are mean±SD for change from baseline to endpoint; P-values are from ANOVA for difference among groups. QUICKI = quantitative insulin sensitivity check index.

As shown in **Supplementary Figure 2**, mean (± SE) post-baseline annual glucose concentrations showed similar patterns for all treatment groups. Mean insulin concentrations showed upward trends in all groups over the course of the study, as would be expected with increasing age, but no significant differences among treatment groups. Interpretation of apparent differences among treatment groups toward the end of the study is hampered by small patient numbers.

Supplementary Figure 2: IGF-I, glucose and insulin by year on study



Thyroid function

Thyroid function tests showed little change, with no significant differences among treatment groups for changes from baseline to endpoint. A summary of key thyroid function parameters is provided below.

Supplementary Table 6: Summary of changes from baseline to endpoint for thyroid function parameters

Group	P/P	P/E	GH/P	GH/E	P-value
TSH (μIU/mL)	-2.01±5.46 (32)	+2.46±23.25 (36)	-0.97±4.78 (33)	+3.57±27.83 (30)	0.569
T4 (μg/dL)	+1.57±2.32 (33)	+1.66±3.00 (36)	+1.48±2.28 (33)	+1.31±2.44 (32)	0.950
Free T4 (ng/dL)	-0.25±0.39 (31)	-0.10±0.40 (37)	-0.18±0.37 (32)	+0.06±1.64 (30)	0.519

Values shown are mean±SD for change from baseline to endpoint; numbers in parentheses are number of patients with values at baseline and endpoint for the analyte. P-values are from ANOVA for difference among groups.

Other laboratory tests

Lipid concentrations, liver function tests, chemistry and hematology showed little change, with no significant differences among treatment groups (data not shown).

Bone age

Mean rate of bone age increase was approximately 1 year per year of study, with no significant differences among groups. **Supplementary Figure 3** presents mean bone age compared with chronological age for the 4 treatment groups.

Supplementary Figure 3: Bone age vs. chronological age