Table of Contents

Otsuka Pharmaceutical Development & Commercialization, Inc.

Investigational Medicinal Product

Brexpiprazole (OPC-34712)

STATISTICAL ANALYSIS PLAN

A Phase 3, 12-Week, Multicenter, Randomized, Double-blind, Placebo-controlled, 2-Arm, Fixed-dose Trial to Evaluate the Efficacy, Safety, and Tolerability of Brexpiprazole (OPC-34712) in the Treatment of Subjects With Agitation Associated With Dementia of the Alzheimer's Type

Protocol No. 331-14-213
IND No. 115,960
EudraCT No. 2017-003940-19

Version: 2.0

Date: 25 January 2021

CONFIDENTIAL - PROPRIETARY INFORMATION

Table of Contents

List of In-text Tables

List of Appendices

List of Abbreviations and Definition of Terms

Abbreviation	Definition
AAD	Agitation in Alzheimer's dementia
AIM	Abnormal involuntary movement
AIMS	Abnormal Involuntary Movement Scale
ALT	Alanine transaminase
ANCOVA	Analysis of covariance
ARH1	Heterogeneous autoregressive of order 1
AST	Aspartate transaminase
BARS	Barnes Akathisia Rating Scale
BMI	Body mass index
bpm	Beats per minute
BUN	Blood urea nitrogen
CGI-I	Clinical Global Impression Improvement Scale
CGI-S	Clinical Global Impression Severity of Illness Scale
CMAI	Cohen-Mansfield Agitation Inventory
CMH	Cochran-Mantel-Haenszel
COVID-19	Coronavirus Disease 2019
CPK	Creatine phosphokinase
CRF	Case report form
CSH	Heterogenous compound symmetry
CST	Clinical Surveillance & Training
DMC	Data Monitoring Committee
ECG	Electrocardiogram
EPS	Extrapyramidal symptoms
ET	Early termination
FDA	Food and Drug Administration
HDL	high-density lipoprotein
IA	Interim Analysis
IAP	Interim Analysis Plan
ICF	Informed consent form
IMP	Investigational medicinal product
ITT	Intent-to-treat
LDL	Low-density lipoprotein
LOCF	Last-observation-carried-forward
LS	Least squares
MAR	Missing at Random
MCAR	Missing Completely at Random
MCMC	Monte Carlo Markov Chain
MedDRA	Medical Dictionary for Regulatory Activities
MI	Multiple imputation
MMRM	Mixed-effect model repeat measurement
MMSE	Mini Mental State Examination
MNAR	Missing Not at Random

Abbreviation	Definition
[redacted]	[redacted]
OC	Observed Case
OTC	Opportunity to Complete
PCR	Potentially clinically relevant
SAP	Statistical analysis plan
SD	Standard deviation
SGOT	Serum glutamic oxaloacetic transaminase
SGPT	Serum glutamic pyruvic transaminase
Sheehan-STS	Sheehan- Suicidality Tracking Scale
TEAE	Treatment-emergent adverse event
TOEPH	Heterogeneous Toeplitz
ULN	Upper limit of normal
UN	Unstructured

1 Introduction

This statistical analysis plan (SAP) expands the statistical section of the protocol 331-14-213 Amendment 3 and documents the statistical methodology and data analysis algorithms and conventions to be applied for statistical analysis and reporting of efficacy, safety and tolerability data of the study. All amendments to the protocol and the addendums to protocol amendment are taken into consideration in developing this SAP.

2 Trial Objectives

Primary: To confirm the efficacy of brexpiprazole compared with placebo in subjects with Agitation in Alzheimer's dementia (AAD)

Secondary: To evaluate the safety and tolerability of brexpiprazole compared with placebo in subjects with AAD

3 Trial Design

This is a phase 3, 12-week, multicenter, randomized, double-blind, placebo-controlled, fixed-dose trial designed to assess the efficacy, safety, and tolerability of brexpiprazole compared with placebo. Subjects will be randomized in a 2:1 ratio to brexpiprazole or placebo. Within the brexpiprazole arm, subjects will be further randomized in a 1:2 ratio to 2 mg/day or 3 mg/day, to explore the efficacy, safety, and tolerability of 2 mg/day and 3 mg/day brexpiprazole versus placebo. The randomization will be stratified by site.

The trial consists of a 2- to 42-day screening period, a 12-week double-blind treatment period, and a 30-day post-treatment safety follow-up period. In addition, for all subjects who terminate early from the trial, attempts will be made to collect data on mortality status by telephone contact with the subject's caregiver at Week 16.

This trial will utilize a two-stage group sequential design with one interim analysis (IA) and one final analysis. One unblinded interim analysis of efficacy data will be performed by an independent Data Monitoring Committee (DMC) on approximately the first 255 randomized subjects who have either completed the Week 12 visit or discontinued early from the trial. The trial may stop for efficacy or futility at the conclusion of the IA or may continue to the final analysis when reaching the maximum planned sample size of about 330 randomized subjects. The sponsor will remain blinded to the observed results of the IA, and will only receive recommendation regarding whether to stop or continue the trial as per the interim analysis plan (IAP) and the DMC Charter.

The trial is organized as follows:

Screening Period

The screening period will range from 2 days up to 42 days, with the goal of completing all screening activities within 30 days, if possible, and will begin when the informed consent form (ICF) is signed, prior to the initiation of any procedures. The screening period may be extended after discussion with and approval by the medical monitor. An eSource method will be used to obtain an identification number for each subject with documented consent.

The purpose of the screening period is to determine the subject's eligibility and to washout prohibited concomitant pharmacotherapy prior to randomization.

External quality oversight methods will be used by Clinical Surveillance & Training (CST) and the Independent Adjudication Panel to promote appropriate subject enrollment.

In addition, starting at screening and continuing throughout the 12-week double-blind treatment period, the subject's behavior will be logged into a diary by the caregiver or facility staff. This diary will be sent to CST on a routine basis in order to corroborate information recorded on the Cohen-Mansfield Agitation Inventory (CMAI).

12-week, Double-blind Treatment Period

Based on a randomization scheme, eligible subjects will be randomized in a 2:1 ratio to one of the following 2 treatment groups:

- Brexpiprazole (further randomized in a 1:2 ratio to 2 mg/day or 3 mg/day)
- Placebo

Subjects will follow a titration schedule, depending upon their assigned treatment group, to gradually increase their dose of the investigational medicinal product (IMP) to their assigned target dose as follows:

[redacted]

The first dose of IMP (brexpiprazole or placebo) will be administered on the day after the baseline visit (ie, Day 1) and ending on Week 12 or early termination (ET; the last day of the treatment period).

Subjects unable to tolerate their assigned dose of brexpiprazole (or matching placebo) will be withdrawn from the trial. Down titration is not allowed at any time during the trial. If a subject is withdrawn or discontinues prematurely before Week 12, every effort will be made to complete all of the Week 12 or ET evaluations prior to administering any additional medications for the treatment of agitation or other prohibited medications.

Subjects will be evaluated at baseline and at Weeks 2, 4, 6, 8, 10, and 12 during the double-blind treatment period. Trial-related efficacy and safety assessments will be performed as outlined in the Schedule of Assessments.

Subjects who complete the 12-week double-blind treatment period of Trial 331-14-213 may be eligible to enter a 12-week open-label extension trial. If this trial is terminated due to overwhelming efficacy from the IA, subjects in this trial at the time may be offered entry into the open-label extension trial if they choose to participate.

Follow-up Period

All subjects with the exception of those entering the optional open-label extension trial, whether they complete the trial or are withdrawn prematurely for any reason, will be followed up for a safety evaluation 30 (+ 2) days after the last dose of IMP during a clinic visit at either the investigator's site or residential facility, as applicable. If a clinic visit is not possible, the subject should be assessed by telephone contact with the subject and a caregiver. For all subjects who terminate early from the trial, all attempts will be made to collect mortality data by telephone contact with the subject's caregiver at Week 16.

4 Sample Size and Power Justification

The initial sample size was calculated based on the treatment effect of 6.5 points with a standard deviation of 16.5 in the change from baseline to Week 12 in the CMAI total score, to achieve 89% power at a 2-sided alpha level of 0.05. The initial sample size was 300 total subjects randomized in a 2:1 ratio to brexpiprazole or placebo. The statistical assumption of the treatment effect of 6.5 points separation (for brexpiprazole versus placebo) is the same as that used for the two completed phase 3 AAD Trials 331-12-283 and 331-12-284.

To account for an empirical dropout rate of 12-13% (based on the two completed phase 3 AAD Trials 331-12-283 and 331-12-284) and a lack of compliance at one identified clinical site, an increment of at least 30 additional subjects will be needed to maintain the target power. The resulting sample size is about 330 total subjects randomized in a 2:1 ratio to brexpiprazole or placebo. Within the brexpiprazole arm, subjects will be further

randomized in a 1:2 ratio to 2 mg/day or 3 mg/day, to explore the efficacy, safety, and tolerability of 2 mg/day and 3 mg/day brexpiprazole versus placebo.

One IA will be conducted when approximately the first 255 subjects had an opportunity to complete (OTC) the 12-week trial. The conservative Bonferroni critical boundary will be used for the two-stage group sequential analysis. Specifically, the alpha allocated to the IA is 0.015 and the alpha left for the final analysis is 0.035, both two-sided. The two-stage group sequential test will attain a power of 87%.

5 Data Sets and Missing Data

5.1 Data Sets for analysis

The following analysis samples are defined for this trial:

Randomized Sample: Consists of all subjects who were randomized into this trial. Subjects are considered randomized when they are assigned a treatment number by the eSource method at the end of Screening Period. A subject receiving trial treatment outside of the eSource will not be considered randomized, but safety will be reported.

Safety Sample: Consists of all subjects who were administered at least one dose of IMP. Subjects will be excluded from this population only if there is documented evidence (ie, number of tablets dispensed = number of tablets returned or no trial drug dispensed) that the subject did not take trial drug. If a subject is dispensed investigational medicinal product (IMP) and is lost to follow-up, he/she will be considered exposed.

Efficacy Sample: The intent-to-treat (ITT) population consists of all subjects in the randomized sample, who took at least 1 dose of IMP and have a baseline and at least one post-baseline evaluation for the CMAI total score.

5.2 Handling of Missing Data

The CMAI is utilized as the primary efficacy assessment of a subject's level of agitated behaviors. The CMAI consists of 29 items all rated on a 1 to 7 scale with 1 being the "best" rating and 7 being the "worst" rating. The CMAI total score is the sum of ratings for all 29 items. The possible total scores are from 29 to 203. The CMAI total score will be unevaluable if less than 24 of the 29 items are recorded. If 24 to 28 of the 29 items are recorded, the total score will be the mean of the recorded items multiplied by 29 and then rounded to the first decimal place.

In addition, CMAI subscale scores for distinct agitation syndromes, also known as CMAI factors of agitation, will be calculated. The three factors include: aggressive behavior, physically nonaggressive behavior, and verbally agitated behavior. Each CMAI factor will be calculated only when all items for the factor are recorded and non-missing. CMAI factors will be derived based on a subset of CMAI items[8] as described in Section 10.3.1.

6 Trial Conduct

6.1 Subject Disposition and Reasons for Discontinuations

Subject disposition is summarized for the randomized sample. Disposition is summarized by treatment group and by subgroup of gender, age (< 65; ≥ 65 and < 75; or ≥ 75), race, and region (North America or Other).

Reasons for discontinuation will be summarized for the randomized sample by treatment group and by subgroup of gender, age, race, and region.

6.2 Treatment Compliance

Compliance in taking investigational medicinal product is calculated by dividing the number of tablets taken by the total number of tablets the subjects were scheduled to take during the trial period. For lost-to-follow-up subjects, last IMP end date record will be used as the treatment end date.

Number (%) of subjects meeting compliance cut-offs (< 70%, ≥ 70% and < 80%, ≥ 80% and < 90%, ≥ 90%) will be summarized by treatment group.

6.3 Protocol Deviation

Protocol deviations are summarized by center and type of deviation for randomized subjects by treatment group. A listing of protocol deviations will also be generated.

7 Baseline Characteristics

7.1 Baseline Definition

For analyses of the double-blind treatment period data, the baseline is the Baseline measurement (expected to be at Day 0). Baseline measurement is defined as the last available measurement prior to the start of double-blind IMP.

7.2 Demographic Characteristics

Baseline demographic characteristics including age, gender, race, ethnicity, height, weight, waist circumference, and body mass index (BMI) will be tabulated by treatment group for all randomized subjects. Additional summaries by the following subgroups will be also generated: by gender, by age group (< 65; ≥ 65 and < 75; or ≥ 75), by race, and by region (North America or Other).

Mean, range and standard deviation will be used to describe continuous variables such as age. Frequency distributions will be tabulated for categorical variables such as race.

7.3 Disease History

A summary of medical, psychiatric, neurological (excluding Alzheimer's), and Alzheimer's disease history will be presented for the Randomized Sample (by treatment group and overall).

7.4 Baseline Disease Characteristics

For the Randomized Sample, baseline and baseline disease characteristics will be summarized by treatment group and overall. The following baseline characteristics will be summarized at baseline: number (%) of institutionalized / non institutionalized subjects; CMAI total score; CMAI derived agitation factors of aggressive behavior, physically nonaggressive behavior, and verbally agitated behavior; Clinical Global Impression-Severity of Illness Scale (CGI-S) score, [redacted] Mini Mental State Examination (MMSE) score; Sheehan Suicidality Tracking Scale (Sheehan-STS) score.

Number of subjects with presence of psychotic symptoms will be summarized at baseline using [redacted]

8 Efficacy Analysis

All efficacy analyses pertaining to the double-blind treatment period will be performed on the Efficacy Sample, and subjects will be included in the treatment group as randomized.

[redacted]

[redacted]

8.1 Primary Efficacy Analysis

The primary efficacy endpoint is the change from the baseline (Day 0 visit) to the end of the double-blind treatment period (Week 12 visit) in CMAI total score. The null hypothesis is that there is no difference in the mean change from baseline to Week 12 in CMAI total score between the two treatments brexpiprazole and placebo.

The primary analysis will be performed on the Efficacy Sample which includes all randomized subjects who took at least one dose of IMP in the double-blind treatment period and who have both a baseline and at least one post-randomization CMAI total score during the double-blind treatment period.

8.1.1 Primary Estimand

In order to support the primary objective of the trial, treatment effect will be estimated under hypothetical situation had no subjects discontinued early from treatment. Due to this strategy, the last collected efficacy assessment after premature trial discontinuation will be done only once at the Early Termination (ET) Visit. Every effort will be made to complete all the ET evaluations prior to administering any additional medications for the treatment of agitation or any other prohibited medications. In the case of terminal or lost to follow-up event, no ET evaluations would be expected, and only scheduled assessments performed before such an event would be used for the primary efficacy analysis. All collected efficacy assessments will be slotted as per the Visit Window Definition, and only one assessment per visit window will be used for the primary efficacy analysis. If there are multiple assessments available in the same window, the latest one will be used.

This trial seeks to clarify the efficacy of brexpiprazole in improving the CMAI total score in the idealistic scenario of full adherence to the assigned treatment in all subjects. In clinical trial practice, however, instances of non-adherence such as treatment discontinuation are likely to occur before the study endpoint (i.e., the Week 12 Visit for this trial). Given this consideration, the hypothetical estimand is the most appropriate for achieving the primary objective of the trial. The primary estimand is described by five attributes as follows.

- Target Population: Subjects with agitation associated with dementia of the Alzheimer's type who have met protocol inclusion/exclusion criteria

- Endpoint: Change from Baseline to Week 12 in the CMAI total score
- Intercurrent Events: Intercurrent events refer to premature treatment discontinuation (i.e., early dropout) prior to Week 12 attributable to adverse events or lack of efficacy or withdrawal of consent/assent or any other causes.
- Measure of Intervention effect: Difference in endpoint means between the brexpiprazole and placebo treatment arm.
- Treatment regimen: 2 or 3 mg daily for 12 weeks.

During the course of this trial, the COVID-19 pandemic broke out. The pandemic had a significant impact on many aspects of clinical trials. There are occasionally virtual visits (i.e., remote assessments) and possibly early discontinuation of treatment directly or indirectly related to the pandemic. However, subjects were required to attend the screening, Day 1 and the end of trial visits in person (i.e., face-to-face). Note that virtual visits in the pandemic environment will not be treated as an intercurrent event for the primary analysis. Subjects who drop out with a reason relating to the COVID-19 pandemic will be handled as they would have dropped out for another reason if the pandemic had not happened.

The hypothetical strategy of handling intercurrent events will be used to clarify the efficacy of the brexipiprazole had there be no occurrence of intercurrent events, regardless of being COVID-19 related or not. In other words, the estimand as described above will use the hypothetical strategy to address the treatment effect of interest that would be envisioned under the hypothetical setting of no occurrence of intercurrent events in the planned 12-week treatment period.

The estimator will be the Mixed Model Repeated Measurements (MMRM) estimate for treatment difference at Week 12, based on all observed case (OC) data until discontinuation from the trial. This reflects the chosen strategies for the identified intercurrent events. Details of the model is provided in the next section.

The OC data consist of actual observations recorded at each visit during the double-blind treatment period. The term “OC data” means that longitudinal data, if missing, will not be imputed by applying an imputation rule (e.g., the LOCF rule). For example, if the subject had a missing CMAI total score at Week 12, then his score at Week 12 will not be imputed with his score observed at an earlier visit. Note that in this SAP the OC data will be used as input data for all the MMRM analyses, though “OC” is not explicitly mentioned.

8.1.2 Method of the Primary Analysis

Mean change from baseline in CMAI total score will be analyzed using a restricted maximum likelihood (REML)-based repeated measures approach. The analysis will be performed by fitting the MMRM model with an unstructured (UN) variance-covariance structure in which the change from the baseline in CMAI total score (Week 2, 4, 6, 8, 10, 12) will be the dependent variable based on the OC data. The analysis will include treatment (brexpiprazole and placebo), trial center, visit week, and an interaction term of treatment by visit week as the fixed, categorical effects and include the interaction term of baseline CMAI total score by visit week as covariates. In case there is a convergence problem with the MMRM model with the UN variance-covariance matrix, the following structures will be used in the order of 1) heterogeneous Toeplitz (TOEPH), 2) heterogeneous autoregressive of order 1 (ARH1), and 3) heterogeneous compound symmetry (CSH) and the first variance-covariance structure converging to the best fit will be used as the primary analysis.

The Kenward-Roger approximation will be used to estimate denominator degree of freedom and adjust standard errors. The contrast (i.e., difference in least-square means between brexpiprazole and placebo) at the Week 12 visit will be estimated from the interaction term of treatment by visit week and will serve as the primary treatment comparison. The point estimate and the 95% confidence interval estimate of the contrast at the Week 12 visit will be reported. Significance test will be based on the contrast estimate at the Week 12 visit by using a two-sided 0.015 level for the interim analysis and a two-sided 0.035 level for the final analysis (when the trial does not stop at the interim analysis). Analysis will be implemented using the software Statistical Analysis System® (SAS® version 9.4 or later) procedure PROC MIXED.

If a structured variance-covariance is used, the empirical sandwich estimator of the standard error of the fixed effects parameters will be used in order to deal with possible model misspecification of the variance-covariance matrix.

Small centers will be defined as centers that do not have at least one evaluable subject (evaluability with regard to the primary efficacy variable) in each treatment arm in the double-blind treatment period. All small centers will be pooled within region to form "pseudo centers" for the purpose of analysis according to the following algorithm. Small centers will be ordered from the largest to the smallest based on the number of evaluable subjects. The process will start by pooling the largest of the small centers with the smallest of the small centers until a non-small center is formed. This process will be repeated using the centers left out of the previous pass. In case of ties in center size, the center with the smallest center code will be selected. If any centers are left out at the end

of this process, they will be pooled with the smallest pseudo centers, or if no pseudo centers exist, they will be pooled with the smallest non-small center. The pooling of the small centers will be performed within each region (North America and Other).

The chosen MMRM analysis assumes missing data to be "missing at random" (MAR), which is a reasonable assumption in longitudinal clinical trials.[2] However, the possibility of "missing not at random" (MNAR) data can never be ruled out. In order to further evaluate robustness of the primary results to the deviations from MAR assumptions, additional sensitivity analysis will be conducted. Sensitivity analyses based on selection model[3], pattern-mixture model[4,5,6], and/or shared parameter model[7] will be performed in order to explore data missing mechanisms of MNAR and investigate the response profile of dropout reason.

In case of gross violations of the MMRM model assumptions, additional supportive analyses will be provided using the generalized version of the Cochran-Mantel-Haenszel (CMH) procedure (van Elteren test), controlling for trial site.

8.1.3 Technical Computation Details for Primary Efficacy Analysis

1) For primary analyses at Week 12 during the double-blind treatment period, the following algorithm will be used for including subjects in the 12-week analysis.
   a) All scheduled visits during the 12 weeks after randomization, regardless of compliance to trial assessment schedule, will be included in the 12-week analysis.
   b) All early terminated (ET) visits before week 12 will be included according to visit window in the 12-week analysis.
   c) All ET visits > 12 weeks after randomization and outside of visit window will not be included in 12-week analysis.
   d) The week number for an ET visit will be calculated by (date of ET visit

[redacted]

Baseline is the CMAI total score at baseline (Day 0 visit).

8.1.4 Sensitivity Analyses for the Primary Efficacy Endpoint

Traditionally the dropout mechanisms are divided into three types[4]: (1) Missing Completely at Random (MCAR), in which the probability of dropout doesn't depend on the observed data and the missing data; (2) MAR, in which the probability of dropout depends on the observed data; and (3) MNAR, where the probability of dropout depends on the missing data and possibly the observed data.

Most of MNAR methods have treated all observations with dropout as if they fall within the same dropout type[3]. In practice, we would find that different dropout reasons may be related to the outcomes in different ways, for example, detailed dropout reasons for this trial are: adverse events (AE), lost to follow-up, protocol deviation, sponsor discontinued trial, subject met (protocol-specified) withdrawal criteria, subject was withdrawn from participation by the investigator, and subject withdrew consent to participate. Dropout due to an AE may lead to MNAR dropout. Subject withdrew consent may also lead to MNAR dropout. However, it is debatable whether a dropout caused by subjects withdrew consent is MAR or MNAR. Except AE, and subject withdrew consent, all other dropout reasons may be assumed as either MCAR or MAR dropout.

As sensitivity analyses for MAR assumption, analyses for MNAR will be carried out. Pattern Mixture Models (PMM) based on Multiple Imputation (MI) with mixed missing data mechanisms will be used to investigate the response profile of dropout subjects by last dropout reason under MNAR mechanism for the following three scenarios:

- Dropout reasons due to AE as MNAR
- Dropout reasons due to either AE or subject withdrew consent as MNAR
- All dropouts as MNAR

Delta Adjustment Imputation Methods

This MNAR sensitivity analysis is to departure from MAR assumption by progressively increasing the delta until conclusion from the primary analysis is overturned. The delta is 0%, 10%, 20%, 30%, ..., 100% of the expected treatment difference of 6.5 points and/or the observed treatment difference between brexpiprazole and placebo from the primary analysis of MMRM model until conclusion of the primary analysis is overturned. When delta = 0 the missing data are assumed to be MAR. When delta > 0, the missing data are assumed to be MNAR.

1) Using Monte Carlo Markov Chain (MCMC) methodology from PROC MI to impute the intermittent missing data to a monotone missing pattern
2) Using a standard MAR-based multiple imputation approach from PROC MI to impute the monotone missingness data
3) For subjects in the treated group and with a dropout reason of AE or subject withdrew consent, a delta will be added for all the values after the dropout time
4) Using MMRM model in the primary analysis to analyze the completed data using PROC MIXED on the multiple imputed data
5) Obtaining the overall results using PROC MIANALYZE.

Placebo Based Imputation Methods

Similar to the "Standard" multiple imputations, except parameters for imputation model obtained from only the placebo (control) group. Missing data for both placebo and drug group are imputed based on the imputation model derived from placebo data. If drug improved outcomes prior to dropout, this benefit is carried into subsequent imputed values, but will diminish over time in accordance with the correlation structure.

8.1.5 Subgroup Analyses for Primary Efficacy Endpoint

Subgroup analyses of change from baseline in CMAI total score to every trial week in the double-blind treatment period will be performed for the following factors:

- Gender
- Race (White and All Other Races)
- Age group (< 65; ≥ 65 and < 75; or ≥ 75)
- Region (North America and Other)

Interaction effects of treatment-by-subgroup will be assessed at Week 12 for the subgroups identified in the previous paragraph. The same MMRM model will be used as for the primary efficacy analysis with the addition of terms for subgroup-by-week and treatment-by-subgroup-by-week. These treatment-by-subgroup interaction analyses will be presented in statistical documentation.

All subgroup analyses will be conducted using the same MMRM analysis as for the primary efficacy analysis except that the fixed class effect terms trial center will not be included in the model.

8.1.6 COVID-19 Pandemic related Sensitivity Analyses

On March 13, 2020, the national emergence concerning the COVID-19 pandemic was announced in the US. The following analyses will be performed on the Efficacy Sample to evaluate the sensitivity of the primary analysis results to the impact of pandemic. The same model (e.g., with the same set of explanatory variables and the response variable) as

that for the primary efficacy analysis will be used for these analyses specified below. Of note, the definition of intercurrent events and the strategy for handling intercurrent events are identical to that for the primary efficacy analysis.

1) An MMRM analysis based on the pre-COVID Efficacy Sample. The pre-COVID Efficacy Sample comprises those subjects in Efficacy Sample who had completed or discontinued the trial before March 13, 2020.
2) An MMRM analysis based on the Efficacy Sample by using the subset of pre-COVID data. The pre-COVID data consist of the data collected before March 13, 2020.
3) An MMRM analysis based on the non-COVID Efficacy Sample. The non-COVID Efficacy Sample comprises those subjects in Efficacy Sample who had no remote assessment of the primary endpoint.
4) An MMRM analysis based on the Efficacy Sample by using the subset of non-COVID data. At the subject level, the non-COVID data refer to the data that exclude his/her first remote assessment of the primary endpoint and all the assessments thereafter.
5) An MMRM analysis based on the Efficacy Sample by using the subset of data that exclude all remote assessments.

If the trial stops early for efficacy at the conclusion of IA, the above COVID-19 related sensitivity analyses will be performed on all available data including the data accrual post IA database lock. The point estimate along with the 95% confidence interval estimate of the treatment contrast will be presented for each of the above sensitivity analyses.

8.1.7 Compliance related Sensitivity Analyses

Two clinical sites were identified by audit as having had non-compliance issues. The primary analysis as described in Section 8.1.2 and the key secondary analysis as described in Section 8.2 will be repeated on all available data but with the exclusion of the data from the two identified sites.

8.1.8 Sensitivity Analyses based on the OTC Subset

The primary analysis as described in Section 8.1.2 and the key secondary analysis as described in Section 8.2 will be repeated on the data of the OTC subset of the Efficacy Sample.

8.2 Key Secondary Analysis

The key secondary efficacy variable is the change from baseline to Week 12 in the CGI-S score, as related to agitation. It will be analyzed by the same statistical methodology specified for the analysis of the primary efficacy variable. In order to control the overall

type I error rate for this key secondary efficacy analysis, a hierarchical testing procedure will be used so that the overall experiment-wise type I error rate is maintained. Thus, if the primary efficacy analysis for the CMAI total score yields a statistically significant result for the comparison of brexpiprazole versus placebo, then the corresponding comparison for this key secondary efficacy variable (CGI-S score) will be tested.

8.3 Secondary Efficacy Analyses

Secondary efficacy variables include the following:

- Change from baseline to Week 12 in CMAI subscale scores (aggressive behavior, physically nonaggressive behavior, verbally agitated behavior)
- Change from baseline in CMAI total score for each trial visit during the double-blind treatment period
- Change from baseline in CGI-S for each trial visit during the double-blind treatment period
- CGI-Improvement (CGI-I) score at each trial visit during the double-blind treatment period
- CMAI response rate at every scheduled trial visit in the double-blind treatment period, where response is defined as ≥ 40%, ≥ 30%, and ≥ 20% reduction in CMAI total score from baseline;
- CMAI Response Rate based on improvement from baseline in agitation status at every scheduled trial visit in the double blind treatment period;
- CGI-I response rate at every scheduled trial visit in the double-blind treatment period, where response is defined as a CGI-I score of 1 or 2 (very much improved or much improved);

Change from baseline will be evaluated using the same MMRM model described in the primary analysis. The CGI-I score will be evaluated by the Cochran–Mantel–Haenszel row mean score differ test (van Elteren) controlling for trial site in last-observation-carried-forward (LOCF) analysis. Response endpoints will be evaluated by the CMH General Association Test controlling for trial site in LOCF analysis.

CMAI subscale scores or factors and agitated status for the first 3 factors will be derived based on a subset of CMAI items[8] as described in Section 10.3.1.

Improvement in agitation status is defined as subject meeting agitation status criteria at baseline on at least one of the 3 subscale scores (aggressive, physically nonaggressive, verbally agitated) and changing the status to non-agitated for the same subscale, while not becoming agitated on another at each visit.

Each factor will be analyzed by the same statistical methodology specified for the analysis of the primary efficacy variable, based on the Efficacy Sample.

[redacted]

8.5 Interim Analysis

An IA of the efficacy data will be performed by an independent DMC when approximately 255 subjects have completed the 12-week trial or discontinued early from the trial. The sponsor will remain blinded to the IA results and will only receive DMC's recommendation as per the IAP and the DMC charter. The DMC may recommend stopping the trial for efficacy or for futility or continuing the trial to the final analysis.

To control the overall Type-I error rate at 0.05 level (two-tailed) in a strong sense, a conservative Bonferroni critical boundary will be used for this two-stage group sequential trial. Specifically, the significance level for the IA is set at 0.015 (2-sided) and the significance level for the final analysis (if the trial does not stop at the IA) is set at 0.035 (2-sided). The IA will be performed as pre-specified in the IAP.

If the DMC recommends stopping the trial for overwhelming efficacy at the conclusion of IA, the sponsor intends to halt the trial in two weeks after receiving DMC's recommendation. The MMRM analysis for the primary endpoint will be repeated on all available data including the data newly accrued after the IA database lock. A significance

level of 0.05 (two-sided) will be used for this additional analysis. All on-going subjects will be given a chance to rollover to the active treatment extension trial Trial 331-201-00182.

8.6 Multiplicity Adjustment

In this SAP, the term endpoint and hypothesis are interchangeably used. The hierarchical testing procedure[11] will be applied to this trial to strongly control the experimental-wise error rate at 5% (two-tailed), wherein the experimental-wise error rate is the probability of making at least one false rejection of the true nulls arising from repeated testing of multiple hypotheses in a group sequential trial. The testing will be conducted on 5 hypotheses of interest in the sequence specified below, respectively at the IA and at the final analysis (if the trial does not stop at the IA), by using pre-specified significance levels. The alpha-spending for the primary endpoint will be applied to the key secondary endpoint and other efficacy endpoints included in the hierarchical testing procedure.

1) The primary efficacy endpoint via the MMRM analysis with an unstructured variance covariance, based on the OC data.

2) The key secondary efficacy endpoint (i.e., Change from baseline to Week 12 in the CGI-S score) via the MMRM analysis with an unstructured variance-covariance, based on the OC data.

3) Change from baseline to Week 12 in the CMAI subscale score (aggressive behavior) via the MMRM analysis with an unstructured variance-covariance, based on the OC data.

4) Change from baseline to Week 12 in the CMAI subscale score (physically nonaggressive behavior) via the MMRM analysis with an unstructured variance-covariance, based on the OC data.

5) Change from baseline to Week 12 in the CMAI subscale score (verbally agitated behavior) via the MMRM analysis with an unstructured variance-covariance, based on the OC data.

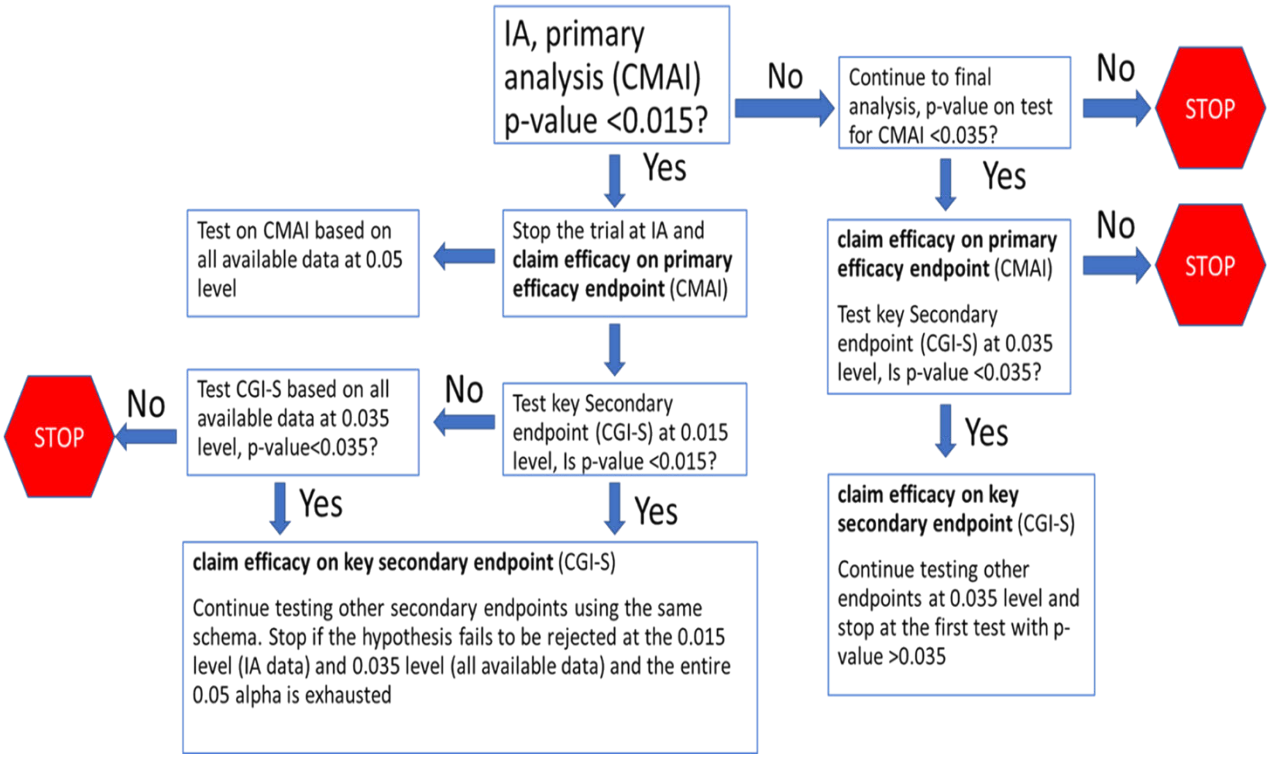
The hierarchical testing procedure is described as follows.

If the trial stops at the IA, the primary efficacy endpoint will be re-tested on all available data (including the recent data accrual post IA database lock) at a two-sided 0.05 level, as mentioned in Section 8.5. Meanwhile, the key secondary endpoint will be tested on the IA data at a two-sided 0.015 level and will also be tested on all available data (including the recent data accrual) at a two-sided 0.035 level. If either of these two tests is nominally

significant at the pre-specified level, then statistical significance on the key secondary endpoint will be claimed and the testing will proceed down to test the CMAI subscale score (aggressive behavior). If neither of these two tests is nominally significant, then statistical significance on the key secondary endpoint cannot be claimed and testing will stop. Testing for each of the other secondary endpoints in the above sequence from 3) to 5) will be conducted similarly to that of key secondary endpoint. Specifically, one test will be conducted on the IA data at a two-sided 0.015 level and the other will be on all available data at a two-sided 0.035 level. Decision rules regarding whether to claim statistical significance and whether to stop the testing will be similar as in the case of testing the key secondary endpoint.

If the trial does not stop at the IA, the above sequence of testing from 1) to 5) will be repeated for the final analysis with the use of a two-sided 0.035 level. In this scenario, the final analysis will be conducted using the data of approximately 330 subjects. There will be no data accrual after the database lock for the final analysis.

Below is the flowchart illustrating the above multiple testing procedure:



For completeness of the presentation, all unadjusted (unadjusted for multiplicity) *p*-values will be reported regardless of whether the endpoints are nominally significant or not.

8.7 Bias Estimation

The naïve point estimate of the parameter mean difference is usually biased upward in a group sequential trial with early efficacy stopping. In this trial, the bias of the estimate of mean difference in change from baseline CMAI total score at Week 12 between brexpiprazole and placebo is fairly small due to (1) the sample size is not small; (2) information fraction for the IA is large, which is 255/330 = 0.77. Regardless of the true mean difference θ, the bias of the mean difference estimate will not exceed 3.02% unit of standard deviation, per the analytic expression for bias (see the penultimate expression on Page 177 of Jennison & Turnbull (2000; Section 8.3)). As a result, the upper bound of the bias will be less than 0.5 point (in terms of the CMAI total score) assuming one unit of within-group standard deviation of 16.5 points.

If the trial stops for efficacy at the IA, a rough bias estimate will be calculated by simply plugging the naïve estimate of mean difference θ into the following expression, which is a simplification from the bias expression mentioned above.

$$\text{bias} = 3.02\% \cdot \sigma \cdot exp\{-(b - \theta\sqrt{I_1})^2/2\}, \text{ where } I_1 = \frac{85}{1.5\sigma^2} \text{ and } b = \Phi^{-1}(1 - 0.015/2) = 2.432$$

and σ is the pooled within-group standard deviation.

The small overestimation bias inherent in the analysis of a group sequential trial is practically negligible given the attenuating circumstance as follows. Due to the enrollment challenges in the US and Western Europe, this study is expected to supplement enrollment with at least 25% of subjects from Eastern Europe and rest of the world (ROW). Based on the results observed in two completed pivotal studies, it is expected that, due to the different healthcare system and different standard of care in Eastern Europe, the effect size in that region would be reduced when compared to that of observed in US, Western Europe and potentially ROW. The oversampling of subjects from Eastern Europe is expected to result in underestimation of the true effect to such a degree that would negate any potential overestimation bias.

9 Safety Analysis

Standard safety variables to be analyzed include AEs, clinical laboratory tests, vital signs, electrocardiograms (ECGs), and physical examination. In addition, data from the following safety scales will be evaluated: MMSE score, assessments of suicidality (e.g, Sheehan-STS), and extrapyramidal symptoms (EPS; e.g, the Simpson-Angus Scale, abnormal involuntary movement scale [AIMS], and Barnes Akathisia Rating Scale [BARS]). Safety analysis will be conducted based on the Safety Sample. In general, baseline of a safety variable is defined as the last observation of the variable before taking the first dose of IMP, unless specified otherwise. Prospectively defined criteria will be used to identify potentially clinically relevant (PCR) abnormal values for clinical laboratory tests, vital signs, ECGs, body weight, and BMI.

9.1 Adverse Events

All AEs will be coded by system organ class and Medical Dictionary for Regulatory Activities (MedDRA) preferred term. The incidence of the following events will be summarized by treatment group:

- Treatment-emergent AEs (TEAEs)
- TEAEs by severity
- TEAEs with an outcome of death
- Serious TEAEs
- TEAEs leading to discontinuation of the IMP

The above summaries will also be prepared for TEAEs potentially causally related to the IMP.

9.2 Clinical Laboratory Tests

9.2.1 Change from Baseline in Lab Tests

Summary statistics for mean and mean change from baseline in the routine clinical laboratory measurements, prolactin concentrations, coagulation parameters (prothrombin time, activated partial thromboplastin time, and international normalized ratio), HbA1c, and thyroid-stimulating hormone will be provided by treatment and by visit.

9.2.2 Potentially Clinically Relevant Values

In addition, the incidence of treatment-emergent PCR values identified using prospectively defined criteria in Appendix 1 for laboratory tests will be summarized by treatment group. A listing of PCRs by subject and by test will be provided.

9.2.3 Potentially Liver Injury Related Laboratory Test

Total bilirubin level will be checked for any subjects with increased alanine transaminase (ALT) or aspartate transaminase (AST) levels greater or equal to three times the upper normal limits (or baseline).

Liver injury related laboratory test results be will summarized for subjects who met following criteria in the Short-term Controlled Trials and Long-term Open-label Trials groups. The corresponding listing will be provided as well.

- AST or ALT ≥ 3 × upper limit of normal (ULN) and
- T_Bili ≥ 2 × ULN

9.2.4 Metabolic Change

In addition to mean change from baseline, the incidence of treatment emergent significant changes in fasting lipids, fasting glucose, and metabolic syndrome will be summarized by treatment group using the following criteria.

Table 9.2.4-1 Criteria for Treatment-Emergent Significant Change in Lipids and Glucose

Lab Parameter	Baseline[a]	Anytime Post Baseline
LDL Direct, Fasting (mg/dL)	Borderline 100 < 160 Normal/Borderline < 160 Normal < 100 Any Value	High ≥ 160 High ≥ 160 Borderline/High ≥ 100 Increased ≥ 30
HDL Cholesterol, Fasting (mg/dL)	Normal ≥ 40 Any Value	Low <40 Decreased ≥ 20
Triglycerides, Fasting (mg/dL)	Normal < 150 Borderline 150 < 200 Normal/Borderline < 200 Normal < 150 Any Value	High 200 < 500 High 200 < 500 High 200 < 500 Borderline/High/Very High ≥ 150 Increased ≥ 50
Glucose Fasting, Serum(mg/dL)	Normal < 100 Impaired 100 < 126 Normal/Impaired < 126 Any Value	High ≥ 126 High ≥ 126 High ≥ 126 Increased ≥ 10

[a] Baseline is calculated from Day 0; if Day 0 is unavailable screen visit will be used

Table 9.2.4-2 Criteria for Treatment-Emergent Metabolic Syndrome		
Description	**Lab Parameter**	**Anytime Post Baseline[a]**
Central Obesity	Waist Circumference	Males: ≥ 102 cm Females: ≥ 88 cm
Dyslipidemia	Triglycerides	≥ 150 mg/dL
	HDL	Males: < 40 mg/dL Females: < 50 mg/dL
Supine Blood Pressure	Systolic	≥ 130 mmHg
	Diastolic	≥ 85 mmHg
Fasting Glucose	Glucose Fasting, Serum	≥ 100 mg/dL

[a] Baseline is calculated from Day 0; if Day 0 is unavailable screen visit will be used

9.3 Physical and Neurological Examination and Vital Signs

Physical and neurological examination findings will be listed by subject.

Summary statistics for change from baseline in vital signs, body weight, and waist circumference will be provided by treatment group.

In addition, the incidence of treatment-emergent PCR values identified using prospectively defined criteria in Appendix 2 for vital signs and body weight will be summarized by treatment group. Listing of PCRs by subject and by test will be provided.

9.4 12-Lead ECG

Summary statistics for change from baseline in ECG parameters will be provided by treatment and by visit.

In addition, the incidence of treatment-emergent PCR values identified using prospectively defined criteria for ECG will be summarized by treatment group. Listing of PCRs by subject and by test will be provided.

For the analysis of QT and QTc, data from three consecutive complexes (representing three consecutive heart beats) will be measured to determine average values. The following QT corrections will be used for reporting purposes in the clinical study report:

- QTcB is the length of the QT interval corrected for heart rate by the Bazett formula: QTcB=QT/(RR)0.5
- QTcF is the length of the QT interval corrected for heart rate by the Fridericia formula: QTcF=QT/(RR)0.33
- QTcN is the length of the QT interval corrected for heart rate by the FDA Neuropharm Division formula: $QTcN = QT/(RR)^{0.37}$

Categorical changes in ECG parameters during the double-blind treatment period will be summarized based on the criteria in Table 9.4-1.

Table 9.4-1 Categorical Change Criteria in QT/QTc Parameters		
Classification	**Category**	**Criteria**
QT	New onset (≥ 450 msec for men or ≥ 470 msec for women)	New onset in QT means a subject who attains a cut off value during treatment period but not at baseline.
QTc[a]	New onset (≥ 450 msec for men or ≥ 470 msec for women)	New onset in QTc means a subject who attains a cut-off value during treatment period but not at baseline.
	New onset (≥ 450 msec for men or ≥ 470 msec for women) and > 10% increase	New onset and > 10% increase in QTc means a subject who attains a cut off value and > 10% increase during treatment period but not at baseline
	New onset (> 500 msec)	New onset (> 500 msec) in QTc means a subject who attains a value > 500 msec during treatment period but not at baseline.
	Increase 30 - 60 msec	Increase from baseline value > 30 and ≤ 60 msec in QTc
	Increase > 60 msec	Increase from baseline value > 60 msec in QTc

[a] QTc categorical change criteria apply to QTcB, QTcF and QTcN.

9.5 Body Weight, Waist Circumference and Body Mass Index

Analyses of body weight, waist circumference and BMI will be performed for the Safety Sample. Body mass index is defined as weight in kilograms divided by the square of height in meters. The mean change from baseline to Week 12 and the last visit in the double-blind treatment period in body weight will be tabulated and analyzed using analysis of covariance (ANCOVA). The ANCOVA models will include the baseline as a covariate and the treatment group as fixed effect.

Percentages of subjects showing significant weight gain (≥ 7% increase in weight), as well as percentages of subjects showing significant weight loss (≥ 7% decrease in weight) from baseline to Week 12 (OC and LOCF) will be analyzed using CMH General Association Test.

9.6 Simpson-Angus Scale, Abnormal Involuntary Movement Scale, and Barnes Akathisia Rating Scale

The mean change from baseline to Week 12/ET in the double-blind treatment period obtained from the Simpson-Angus Scale total score, AIMS total score (total of the first 7 item scores), and BARS Global Clinical Assessment will be performed for the Safety Sample and tabulated and analyzed using ANCOVA.

The ANCOVA model will include the baseline measure and the treatment group.

The same analyses will be performed on the AIMS individual item scores 8, 9, and 10. In addition, incidence of BARS Global Clinical Assessment of Akathisia during the double-blind treatment period by severity category will be provided. Analyses of these EPS rating scales will be performed for the Safety Sample.

9.7 Mini-Mental State Examination

The mean changes from baseline to Week 12/ET visit in the double-blind treatment period in MMSE will be tabulated and analyzed by treatment group using ANCOVA. The ANCOVA model will include the baseline item score as covariate and treatment group as main effect. The analyses will be performed for Safety Sample.

9.8 Suicidality Data

The mean changes from baseline to Week 12/ET visit in the double-blind treatment period in Sheehan-STS individual item scores, suicidal ideation subscale score, suicidal behavior subscale score, or total score will be tabulated and analyzed by treatment group using ANCOVA. The ANCOVA model will include the baseline item score as covariate and treatment group as main effects.

Incidence of treatment emergent suicidal ideation, suicidal behavior will be summarized by treatment, and overall.

9.9 Concomitant Medications

Number and proportion of subjects taking concomitant medications prior to IMP, during the double-blind treatment period, and after IMP are tabulated by drug classification using the World Health Organization drug dictionary.

9.10 Extent of Exposure

The start date of double-blind IMP with brexpiprazole or placebo will be the first day of double-blind dosing. The number and percentage of subjects who receive double-blind IMP will be presented by week and by treatment group. Each dosing week will be based on the actual week; i.e., Day 1 to 7 in Week 1, Day 8 to 14 in Week 2, etc. This summary will be performed on the Safety Sample.

The number and percentage of completers will be presented by week and by treatment group.

The mean daily dosage will be summarized by week and treatment group using descriptive statistics. The mean daily dosage per subject per week will be determined for each week of the trial. This will be calculated by dividing the sum of individual total doses by the number of days in the week interval. The summary will contain for each

treatment group the number of subjects receiving double-blind IMP, and the mean and range of the mean daily dose for each week.

10 Conventions

10.1 Trial Visit Windows

Trial visit windows will be used to map visits using trial day intervals. This visit window convention applies to tables and listings for all efficacy and safety scales. This derived trial window variable will be named as WEEK and will be footnoted. In listings it will be listed along with the case report form (CRF) trial visit.

[redacted]

10.2 Pooling of small centers

Primary efficacy analysis will be performed on the ITT Efficacy Sample which comprises those subjects in the randomized sample who have a baseline value and at least one post-randomization value for CMAI total score in the double-blind treatment period.

Small centers will be defined as centers that do not have at least one evaluable subject (evaluable with regard to the primary efficacy variable) in each treatment arm in the double-blind treatment period. All small centers will be pooled to form "pseudo centers" for the purpose of analysis according to the following algorithm. Small centers will be ordered from the largest to the smallest based on the number of evaluable subjects (ie, subjects who have a baseline value and at least one post-randomization value for CMAI total score in the double-blind treatment period). The process will start by pooling the largest of the small centers with the smallest of the small centers until a non-small center is formed. This process will be repeated using the centers left out of the previous pass. In case of ties in center size, the center with the smallest center code will be selected. If any centers are left out at the end of this process, they will be pooled with the smallest pseudo centers, or if no pseudo centers exist, they will be pooled with the smallest non-small center. The pooling of the small centers will be performed within each region (North America and Other).

10.3 Scales: Rules for Scoring and Handling of Missing Data

10.3.1 Cohen-Mansfield Agitation Inventory

The CMAI consists of 29 items all rated on a 1 to 7 scale with 1 being the "best" rating and 7 being the "worst" rating. The CMAI total score is the sum of ratings for all 29 items. The possible total scores are from 29 to 203. The CMAI total score will be unevaluable if less than 24 of the 29 items are recorded. If 24 to 28 of the 29 items are recorded, the total score will be the mean of the recorded items multiplied by 29 and then rounded to the first decimal place.

Factor 1: Aggressive behavior: Hitting (including self), kicking, scratching, grabbing onto people, pushing, hurt self or other (cigarette, hot water, etc.), throwing things, cursing or verbal aggression, spitting (include at meals), tearing things or destroying property, screaming, biting.

Criteria for agitated status based on Factor 1:

- at least one aggressive behavior occurring at a frequency of at least 4; or
- at least two aggressive behavior occurring at a frequency of at least 3; or
- at least three aggressive behavior occurring at a frequency of at least 2;

Factor 2: Physically nonaggressive behavior: Pace, aimless wandering, trying to get to a different place, general restlessness, inappropriate dress or disrobing, handling things inappropriately, performing repetitious mannerisms.

Criteria for agitated status based on Factor 2:

- at least one physically nonaggressive behavior occurring at a frequency of at least 5; or

- at least two physically nonaggressive behavior occurring at a frequency of at least 4; or
- at least three physically nonaggressive behavior occurring at a frequency of at least 3; or
- at least four physically nonaggressive behavior occurring at a frequency of at least 2;

Factor 3: Verbally agitated behavior: Complaining, constant unwarranted request for attention or help, repetitious sentences or questions, negativism.

Criteria for agitated status based on Factor 3:

- at least one verbally agitated behavior occurring at a frequency of at least 5; or
- at least two verbally agitated behavior occurring at a frequency of at least 4; or
- at least three verbally agitated behavior occurring at a frequency of at least 3; or
- at least four verbally agitated behavior occurring at a frequency of at least 2;

Factor 4: Hiding things, hoarding things.

The following items are not included in factors derivation[8]: Intentional Falling; Making Verbal Sexual Advances; Making Physical Sexual Advances; Strange Noises (weird laughter or crying); Eating or Drinking Inappropriate Substances.

10.3.2 Clinical Global Impression-Severity of Illness Scale

The severity of agitation for each subject will be rated using the Clinical Global Impression-Severity of Illness Scale (CGI-S). To perform this assessment, the investigator (or designee) will answer the following question: "Considering your total clinical experience with this particular population, how mentally ill (as related to agitation) is the subject at this time?" Response choices are 0 = not assessed; 1 = normal, not at all ill; 2 = borderline mentally ill; 3 = mildly ill; 4 = moderately ill; 5 = markedly ill; 6 = severely ill; and 7 = among the most extremely ill subjects. The score 0 (not assessed) will be set to missing. The CGI-S is therefore a 7-point scale from 1 through 7.

10.3.3 Clinical Global Impression-Improvement Scale

The efficacy of brexpiprazole in the treatment of agitation will be rated for each subject using the CGI-I. The investigator (or designee) will rate the subject's total improvement (as related to agitation) whether or not it is due entirely to IMP. Response choices are 0 = not assessed, 1 = very much improved, 2 = much improved, 3 = minimally improved, 4 = no change, 5 = minimally worse, 6 = much worse, and 7 = very much worse.

This is also a 7-point scale (1 to 7), with 1 being very much improved and 7 being very much worse. The scale also includes 0: not assessed, which will be set to missing for

purposes of analysis. At each visit other than randomization, the global improvement will be judged with respect to subject's condition at randomization.

10.3.5 Hachinski Ischemic Scale (Rosen Modification)

The Rosen-modified Hachinski Ischemic Scale assesses whether a subject's dementia is likely due to vascular causes by the response to 8 questions: abrupt onset, stepwise deterioration, somatic complaints, emotional incontinence, history of hypertension, history of stroke, focal neurologic signs, and focal neurologic symptoms. The Rosen modified Hachinski Ischemic Scale will be completed to assess eligibility for the trial by the same neurologist who performs the neurological examination.

10.3.6 Simpson-Angus Scale

The Simpson-Angus Scale is a rating scale used to measure EPS. The Simpson-Angus Scale is a 10-item scale, with each item rated from 0 to 4, with 0 being normal and 4 being the worst. The Simpson-Angus Scale total score is the sum of ratings for all 10 items, with possible total scores from 0 to 40. The Simpson-Angus Scale total score

will be un-evaluable if less than 8 of the 10 items are recorded. If 8 or 9 of the 10 items are recorded, the total score will be the mean of the recorded items multiplied by 10 and then rounded to the first decimal place.

10.3.7 Abnormal Involuntary Movement Scale

The AIMS is a 12-item scale. The first 10 items are rated from 0 to 4 (0 = best, 4 = worst). An item score of 0, depending on the item, either means: no abnormal involuntary movement (AIM), or no incapacitation due to AIM, or no awareness of AIM. An item score of 4 either means: severe AIM, or severe incapacitation due to AIM, or being aware of, and severe distress caused by AIM. Items 11 and 12, related to dental status, have dichotomous responses, 0 = no and 1 = yes. The AIMS total score is the sum of the ratings for the first seven items. The possible total scores are from 0 to 28. The AIMS total score will be un-evaluable if less than 6 of the first 7 items are recorded. If 6 of the items are recorded, then the total score will be the mean of the recorded items multiplied by 7 and then rounded to the first decimal place.

10.3.8 Barnes Akathisia Rating Scale

BARS will be used to assess the presence and severity of akathisia. This scale consists of 4 items. Only the 4th item, the Global Clinical Assessment of Akathisia, will be evaluated in this trial. This item is rated on a 6-point scale, with 0 being best (absent) and 5 being worst (severe akathisia).

10.3.9 Sheehan Suicidality Tracking Scale

Suicidality will be monitored during the trial using the Sheehan-STS. The Sheehan-STS is a prospective scale that assesses treatment-emergent suicidal thoughts and behaviors. Each item of the Sheehan-STS is scored on a 5-point Likert scale (0 = not at all; 1 = a little; 2 = moderate; 3 = very; and 4 = extremely). The Sheehan-STS can be analyzed as individual item scores, suicidal ideation subscale score, suicidal behavior subscale score, or total score. The trial will use the "Screening" and "Since Last Visit" version of the scale. The "Screening" Sheehan-STS form will be completed at the screening visit to determine eligibility. Any subject with evidence of serious risk of suicide based on the Sheehan-STS, ie, a score of 3 or 4 on any one question 2 through 6 or 11 or a score of 2 or higher on any one questions 1a, 7 through 10, or 12, or who, in the opinion of the investigator, present a serious risk of suicide should be excluded from the trial. The "Since Last Visit" Sheehan-STS form will be completed at all other visits. The medical monitor should be contacted if a score of 3 or 4 on any one question 3 through 6 or 11 or a score of 2 or higher on any one questions 1a, 7 through 10, or 12, or 18, or if suicide results in death.

10.3.10 Mini-Mental State Examination

The MMSE is a brief practical test for assessing cognitive dysfunction. The test consists of 5 sections (orientation, registration, attention and calculation, recall, and language) and has a total possible score of 30. The MMSE is used for screening subjects and is also to be completed at Week 12/ET.

The MMSE is a 19-item scale. Items 1 to 10, 15, and 17 to 19 are rated on a scale from 0 to 1, item 14 is rated on a 0 to 2 scale, items 11, 13 and 16 are rated on a scale from 0 to 3, and Item 12 is rated on a scale from 0 to 5. Low scores are the worst, high scores are the best. The MMSE total score is calculated by adding the individual item scores. The possible range for the MMSE total score is from 0 to 30. If the maximum total of the missing items could contribute more than 6 points to the total score then the total score will be set to missing. Otherwise a mean non-missing items score will be calculated by summing the non-missing items and dividing them by the maximum score possible from the non-missing items. For missing items with possible scores from 0 to 1, the mean score will be imputed for each missing item. If item 14 is missing, two times this mean will be imputed for item 14. If items 11, 13, or 16 are missing three times this mean for will be imputed each missing item. If item 12 is missing 5 times then this mean will be imputed for item 12. After all by-item imputation has been done, the individual item scores will be added and this sum will be rounded to the first decimal place to arrive at an imputed total score. In other terms, the MMSE total score is simply the mean non-missing items score multiplied by 30, and then rounded to the first decimal place.

11 References

1 Jennison C, Turnbull B.W. Group Sequential Methods with Applications to Clinical Trials. 2000; Chapman & Hall/CRC.

2 Siddiqui O, Hung JHM, O'Neill R. MMRM vs. LOCF: A comprehensive comparison based on simulation study and 25 NDA datasets. J Biopharmaceutical Stats. 2009; 19(2):227-46.

3 Diggle P, Kenward MG. Informative drop-out in longitudinal data analysis. Applied Statistics. 1994;43:49-93.

4 Little RJA. Modeling the drop-out mechanism in repeated measures studies. J Am Stat Assoc. 1995;90:1112-21.

5 Hedeker D, Gibbons RD. Application of random effects pattern-mixture models for missing data in longitudinal studies. Psychological Methods. 1997;2:64-78.

6 Ali MW, Siddiqui O. Multiple imputation compared with some information dropout procedures in the estimation and comparison of rates of change in longitudinal clinical trials with dropouts. J Biopharmaceutical Stats. 2000;10(2):165-81.

7 Wu MC, Bailey KR. Estimation and comparison of changes in the presence of informative right censoring: Conditional linear model. Biometrics. 1989;45:939-55.

8 Rabinowitz J, Davidson M, Paul De Deyn P, Katz I, Brodaty H, Cohen-Mansfield J. Factor Analysis of the Cohen-Mansfield Agitation Inventory in Three Large Samples of Nursing Home Patients With Dementia and Behavioral Disturbance. Am J Geriatr Psychiatry. 2005;13(11):991-8.

9 Clinical Protocol, OPDC Protocol No: 331-14-213, Amendment 3. A Phase 3, 12-Week, Multicenter, Randomized, Double-blind, Placebo-controlled, 2-arm, Fixed Dose Trial to Evaluate the Efficacy, Safety and Tolerability of Brexpiprazole (OPC-34712) in the Treatment of Subjects with Agitation Associated with Dementia of the Alzheimer's Type. 14 September 2020.

10 COVID-19 Addendum for Clinical Protocol, OPDC Protocol No: 331-14-213, version 2.1. 06 August 2020.

11 Glimm E, Maurer W, Bretz F. Hierarchical Testing of Multiple Endpoints in Group-sequential Trials. Stat Med. 2010; 29(2):219-28

[redacted]

Appendix 1 Criteria for Identifying Laboratory Values of Potential Clinical Relevance

Laboratory Tests	Criteria
Chemistry	
AST (SGOT)	≥ 3 × ULN
ALT (SGPT)	≥ 3 × ULN
Alkaline phosphatase	≥ 3 × ULN
LDH	≥ 3 × ULN
BUN	≥ 30 mg/dL
Creatinine	≥ 2.0 mg/dL
Uric Acid	
Men	≥ 10.5 mg/dL
Women	≥ 8.5 mg/dL
Bilirubin (total)	≥ 2.0 mg/dL
CPK	≥ 3 × ULN
Prolactin	> ULN
Hematology	
Hematocrit	
Men	≤ 37% and decrease of ≥ 3 percentage points from Baseline
Women	≤ 32% and decrease of ≥ 3 percentage points from Baseline
Hemoglobin	
Men	≤ 11.5 g/dL
Women	≤ 9.5 g/dL
White blood count	≤ 2,800/ mm^3 or ≥ 16,000/ mm^3
Eosinophils	≥ 10%
Neutrophils	≤ 15%
Absolute neutrophil count	≤ 1,000/mm^3
Platelet count	≤ 75,000/mm^3 or ≥ 700,000/ mm^3
Urinalysis	
Protein	Increase of ≥ 2 units
Glucose	Increase of ≥ 2 units
Casts	Increase of ≥ 2 units
Additional Criteria	
Chloride	≤ 90 mEq/L or ≥ 118 mEq/L
Potassium	≤ 2.5 mEq/L or ≥ 6.5 mEq/L
Sodium	≤ 126 mEq/L or ≥ 156 mEq/L
Calcium	≤ 8.2 mg/dL or ≥ 12 mg/dL
Glucose	
Fasting	≥ 100 mg/dL
Non-Fasting	≥ 200 mg/dL
Total Cholesterol, Fasting	≥ 240 mg/dL
LDL Cholesterol, Fasting	≥ 160 mg/dL
HDL Cholesterol, Fasting	
Men	< 40 mg/dL
Women	< 50 mg/dL
Triglycerides, Fasting	≥ 150 mg/dL

Appendix 2 Criteria for Identifying Vital Signs of Potential Clinical Relevance

Variable	Criterion Value[a]	Change Relative to Baseline[a]
Heart Rate[b]	> 120 bpm	≥ 15 bpm increase
	< 50 bpm	≥ 15 bpm decrease
Systolic Blood Pressure[b]	> 180 mmHg	≥ 20 mmHg increase
	< 90 mmHg	≥ 20 mmHg decrease
Diastolic Blood Pressure[b]	> 105 mmHg	≥ 15 mmHg increase
	< 50 mmHg	≥ 15 mmHg decrease
Orthostatic Hypotension	≥ 20 mmHg decrease in systolic blood pressure and a ≥ 25 bpm increase in heart rate from supine to sitting/standing	Not Applicable (baseline status not considered)
Weight	—	≥ 7% increase ≥ 7% decrease

[a] In order to be identified as potentially clinically relevant, an on-treatment value must meet the "Criterion Value" and also represent a change from the subject's baseline value of at least the magnitude shown in the "Change Relative to Baseline" column.

[b] As defined in "Supplementary Suggestions for Preparing an Integrated Summary of Safety Information in an Original NDA Submission and for Organizing Information in Periodic Safety Updates," FDA Division of Neuropharmacological Drug Products draft (2/27/87).

Appendix 3 Criteria for Identifying ECG Measurements of Potential Clinical Relevance

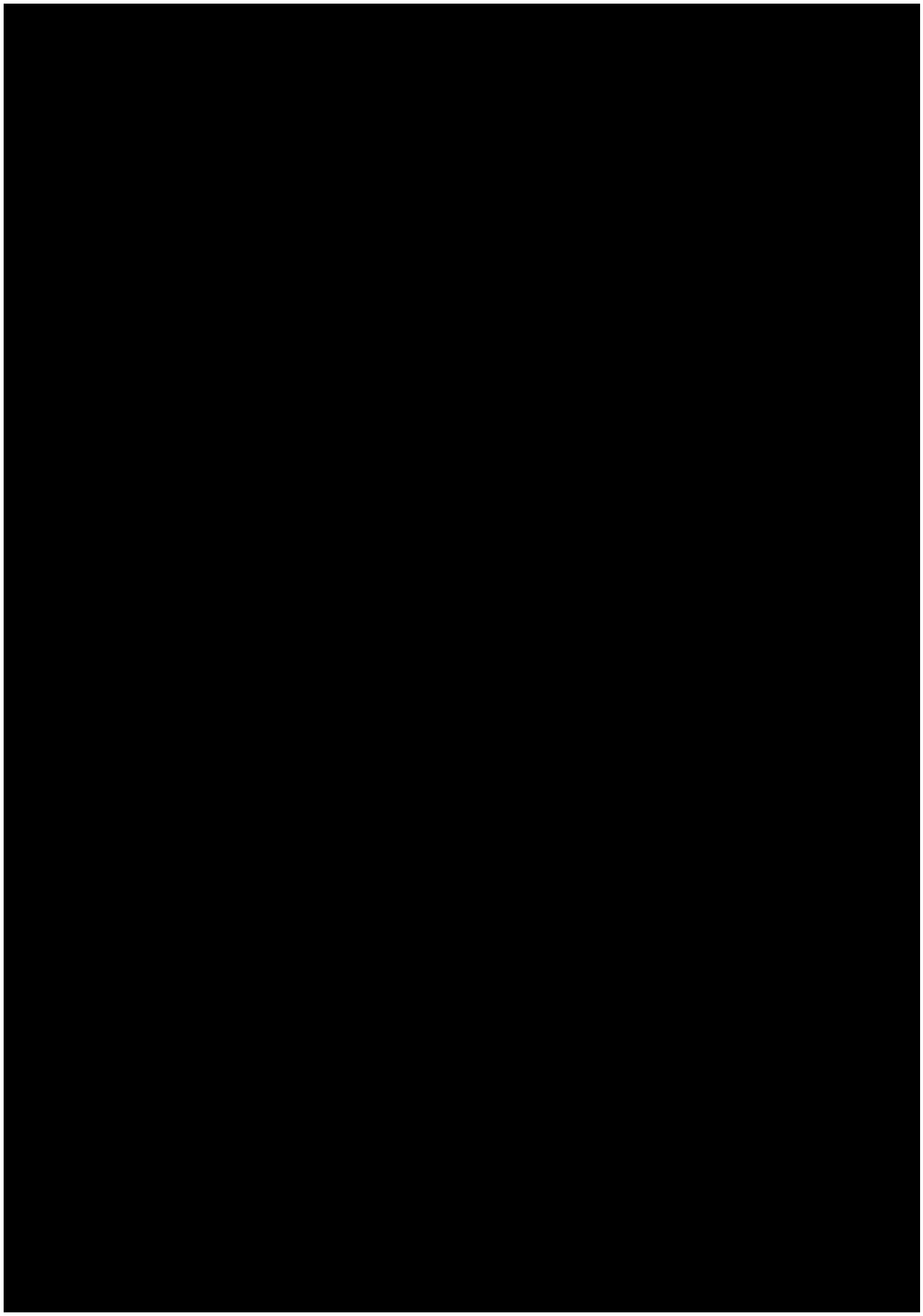
Variable	Criterion Value[a]	Change Relative to Baseline[a]
Rate		
Tachycardia	≥ 120 bpm	increase of ≥ 15 bpm
Bradycardia	≤ 50 bpm	decrease of ≥ 15 bpm
Rhythm		
Sinus tachycardia[b]	≥ 120 bpm	increase of ≥ 15 bpm
Sinus bradycardia[c]	≤ 50 bpm	decrease of ≥ 15 bpm
Supraventricular premature beat	all	not present → present
Ventricular premature beat	all	not present → present
Supraventricular tachycardia	all	not present → present
Ventricular tachycardia	all	not present → present
Atrial fibrillation	all	not present → present
Atrial flutter	all	not present → present
Conduction		
1° atrioventricular block	PR ≥ 200 msec	increase of ≥ 50 msec
2° atrioventricular block	all	not present → present
3° atrioventricular block	all	not present → present
Left bundle-branch block	all	not present → present
Right bundle-branch block	all	not present → present
Pre-excitation syndrome	all	not present → present
Other intraventricular conduction block[d]	QRS ≥ 120 msec	increase of ≥ 20 msec
Infarction		
Acute or subacute	all	not present → present
Old	all	not present → present ≥ 12 weeks post trial entry
ST/T Morphological		
Myocardial Ischemia	all	not present → present
Symmetrical T-wave inversion	all	not present → present
Increase in QTc	QTcF ≥ 450 msec (men) QTcF ≥ 470 msec (women)	

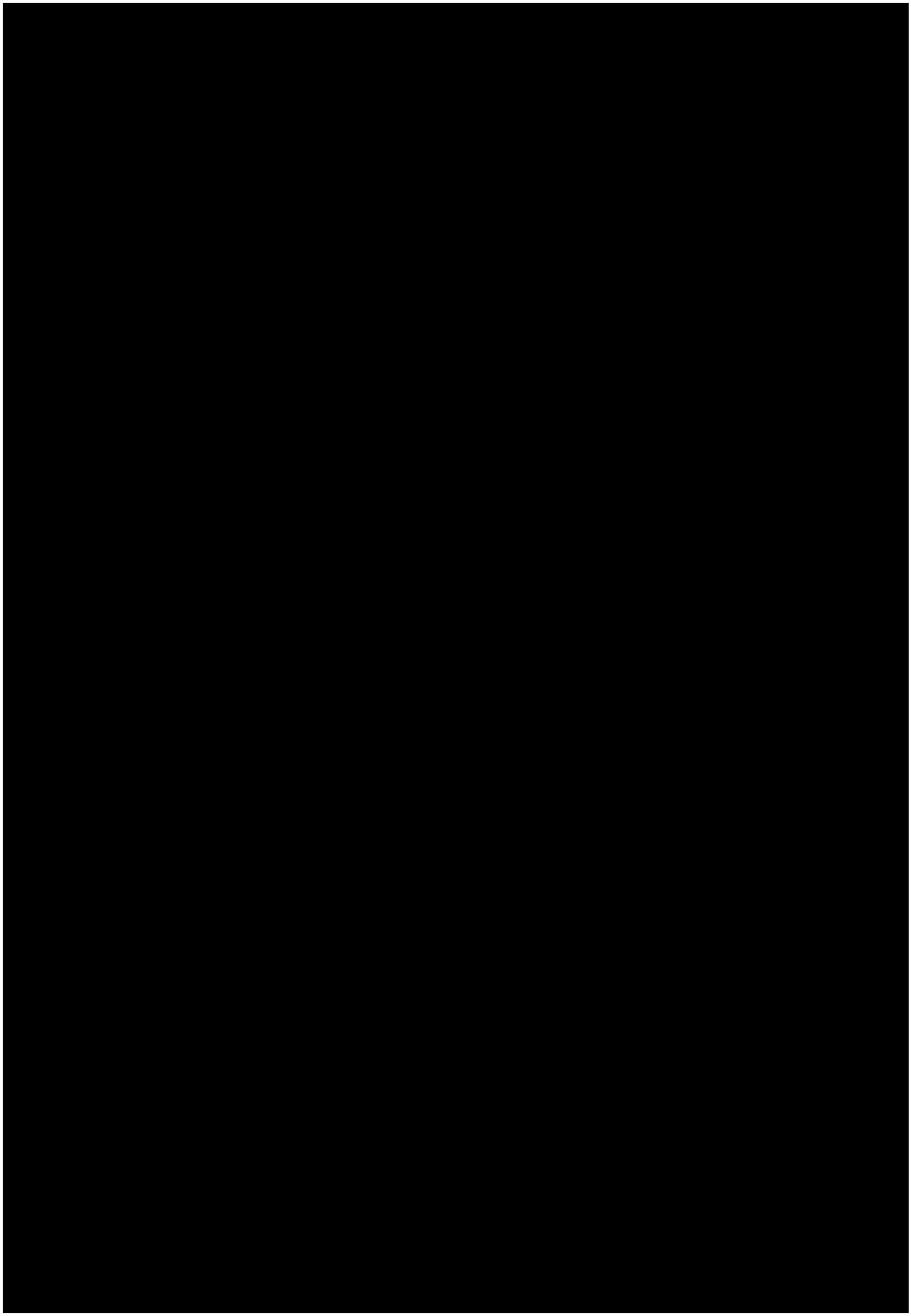
[a] In order to be identified as potentially clinically relevant, an on-treatment value must meet the "Criterion Value" and also represent a change from the subject's baseline value of at least the magnitude shown in the "Change Relative to Baseline" column.

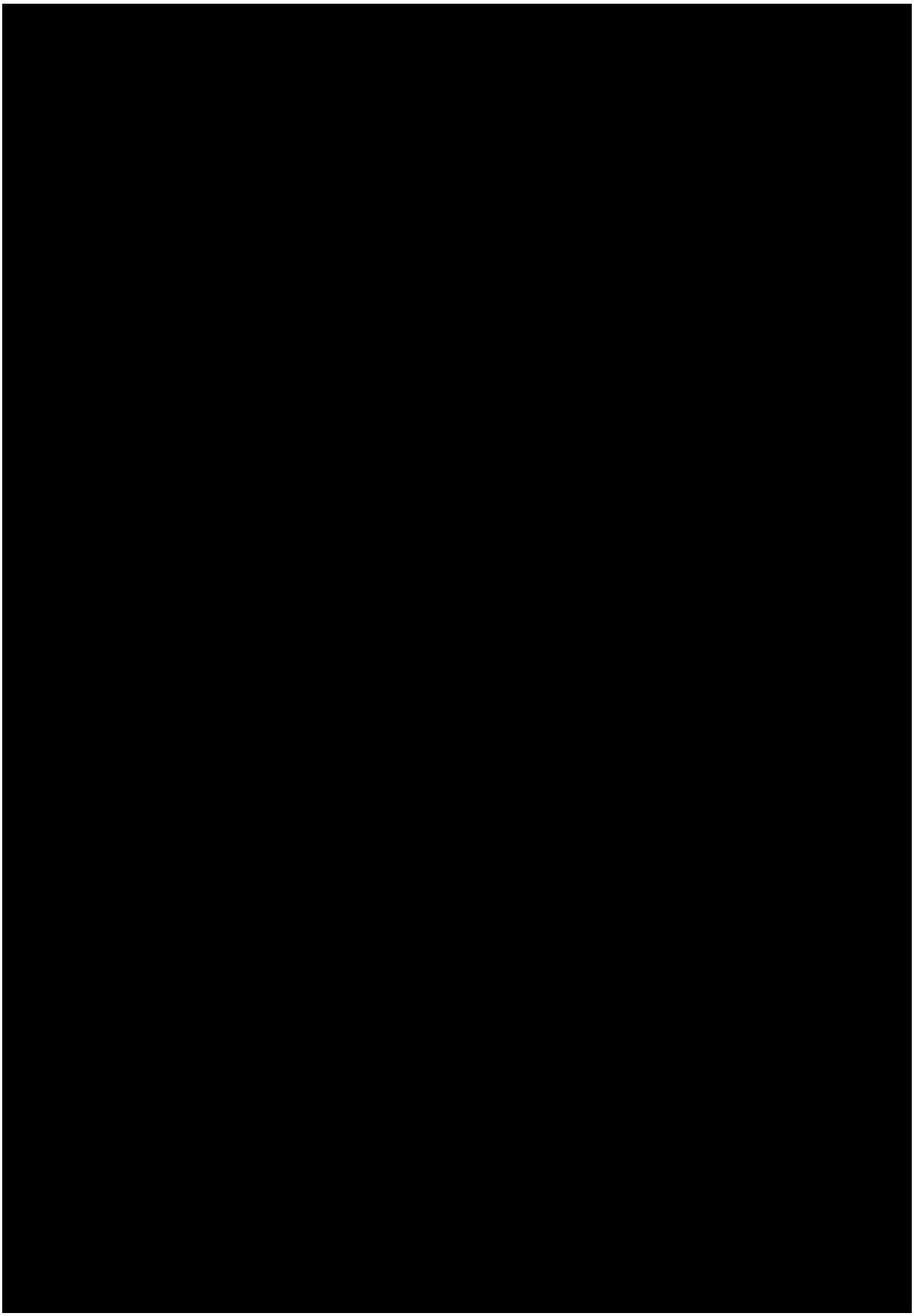
[b] No current diagnosis of supraventricular tachycardia, ventricular tachycardia, atrial fibrillation, atrial flutter, or other rhythm abnormality.

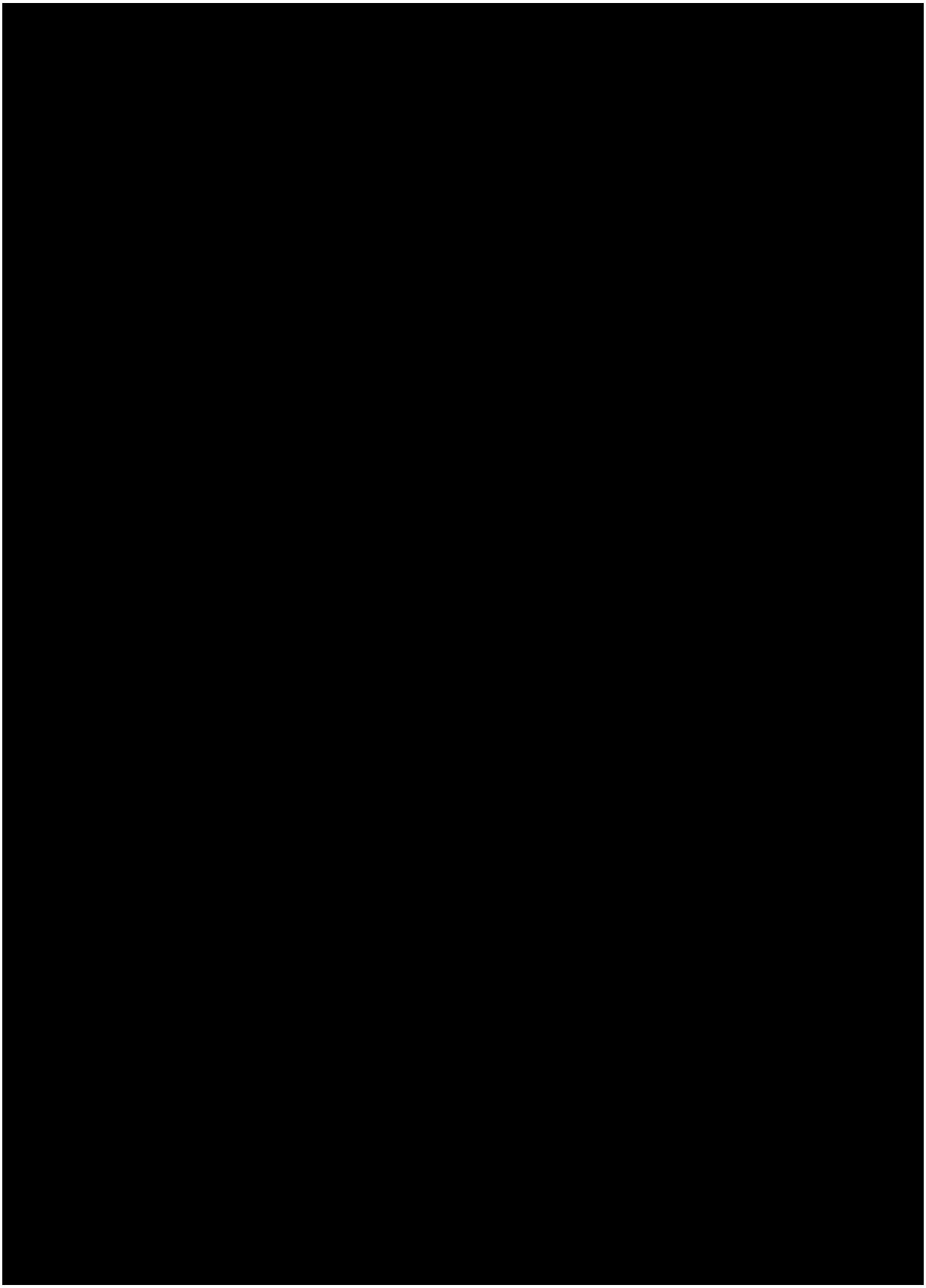
[c] No current diagnosis of atrial fibrillation, atrial flutter, or other rhythm abnormality.

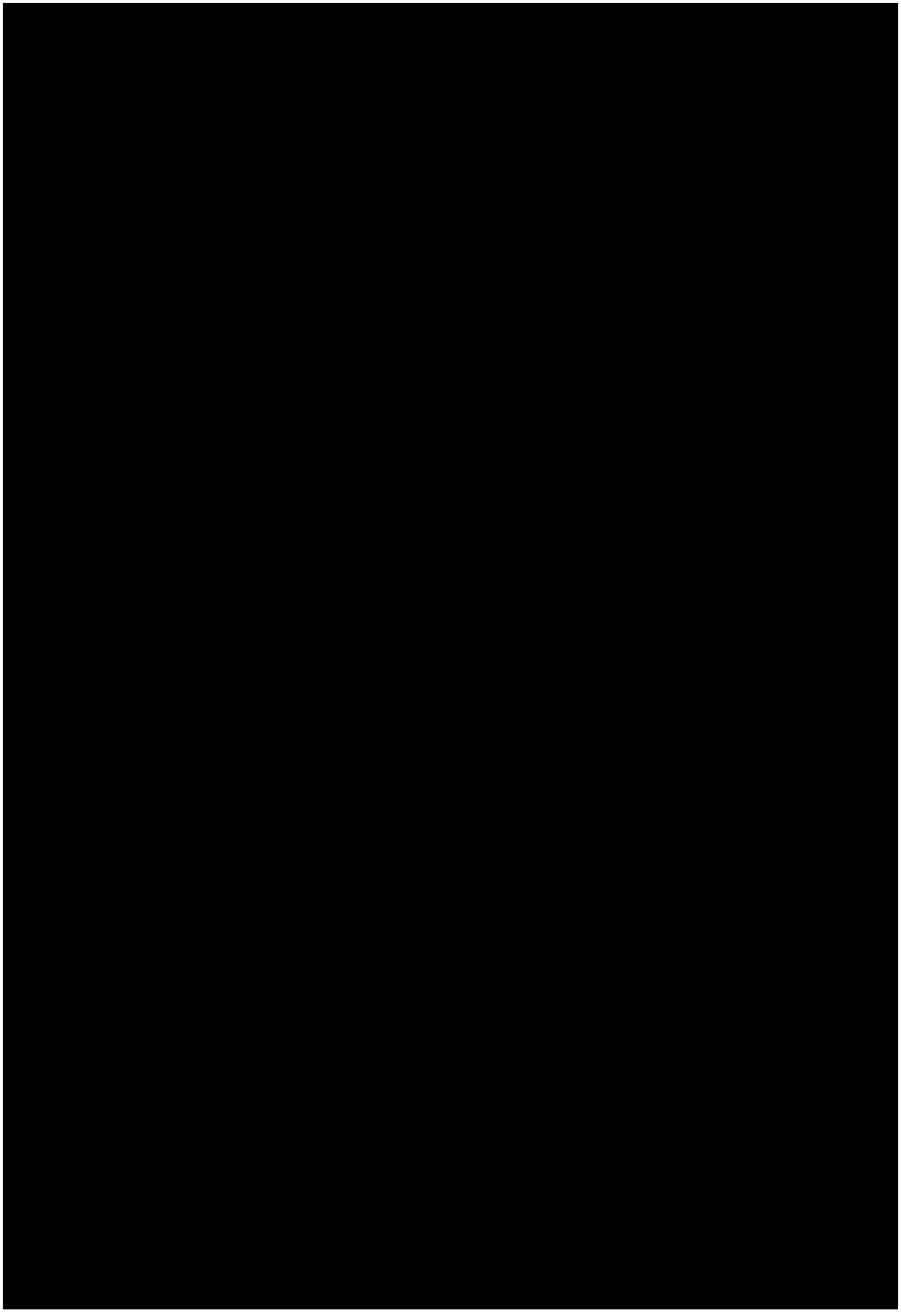
[d] No current diagnosis of left bundle branch block or right bundle branch block.

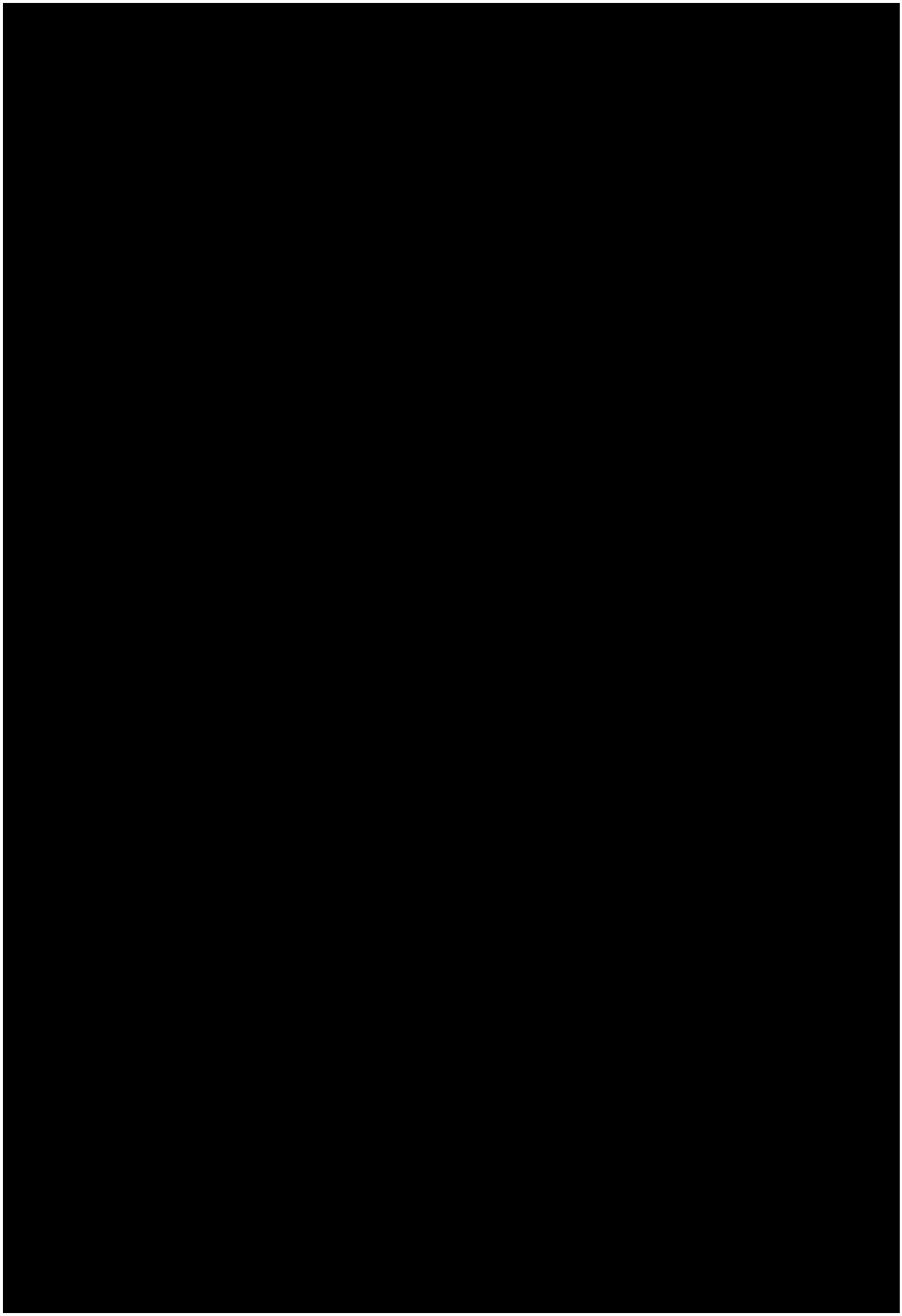


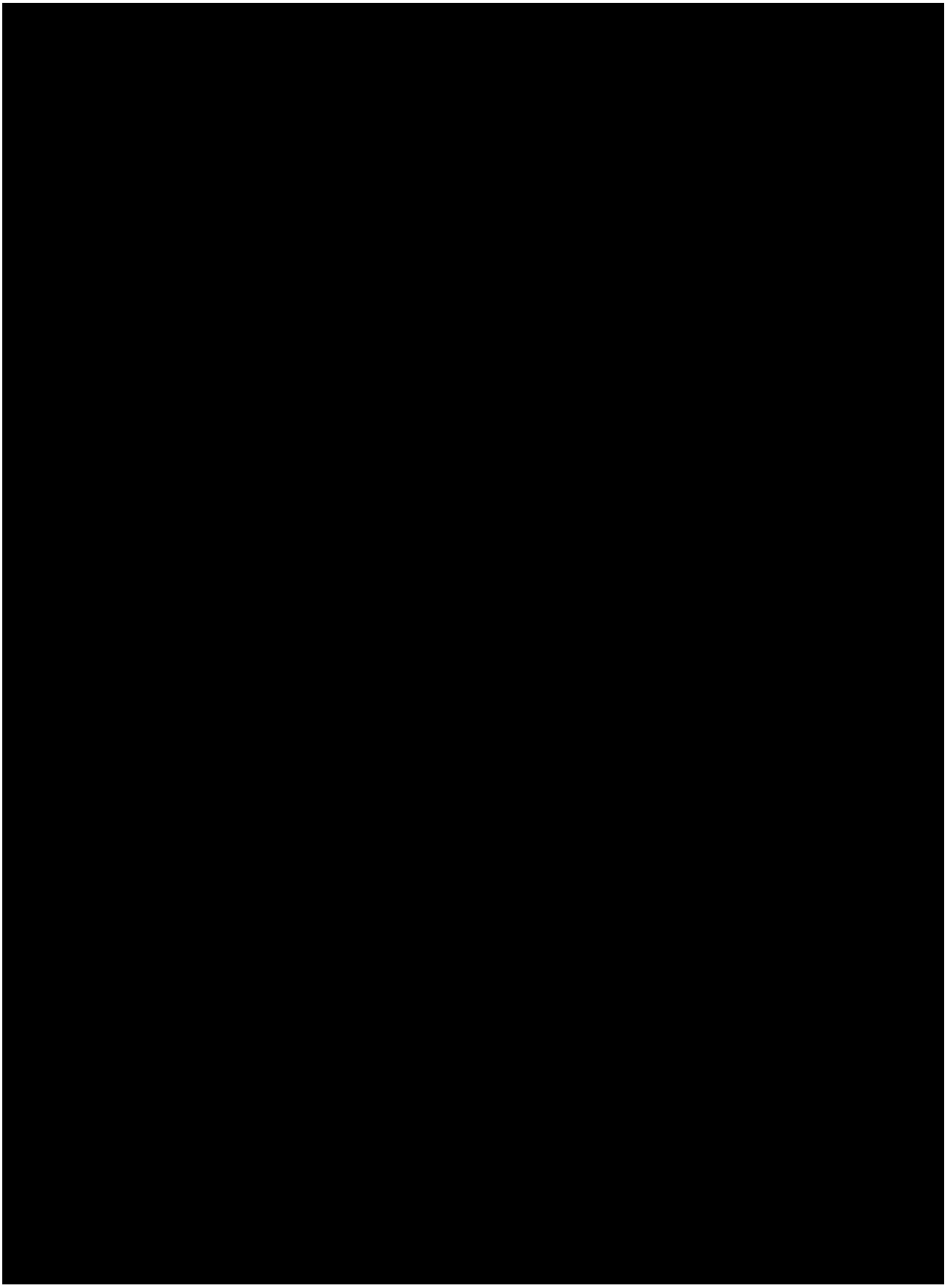












This page is a manifestation of an electronically captured signature

SIGNATURE PAGE

Document Name: P331-14-213_SAP v2.0 _28Jan2021

Document Number: [redacted]

Document Version: 2.0

Signed by	Meaning of Signature	Server Date (dd-MMM-yyyy hh:min) - UTC timezone
[redacted]	Biostatistics Approval	29-Jan-2021 01:56:04
[redacted]	Clinical Approval	28-Jan-2021 20:44:19

Otsuka Pharmaceutical Development & Commercialization, Inc.

Investigational Medicinal Product

Brexpiprazole (OPC-34712)

Addendum to Statistical Analysis Plan Version 2.0

A Phase 3, 12-Week, Multicenter, Randomized, Double-blind, Placebo-controlled, 2-Arm, Fixed-dose Trial to Evaluate the Efficacy, Safety, and Tolerability of Brexpiprazole (OPC-34712) in the Treatment of Subjects With Agitation Associated With Dementia of the Alzheimer's Type

Protocol No. 331-14-213
IND No. 115,960
EudraCT No. 2017-003940-19

Version: 1.0

Date: 08 September 2021

CONFIDENTIAL - PROPRIETARY INFORMATION

Table of Contents

1 Introduction

This addendum to the statistical analysis plan (SAP) version 2.0 is to document a modification to the multiplicity adjustment after the approval of the original SAP (version 2.0, issued on 29 January 29, 2021) and the interim analysis plan (version 4.0, issued on September 10, 2020), and to address FDA's feedbacks on the SAP. The modification will have no impact on the familywise Type-I error rate, which is strongly controlled at a two-sided 5% level. The statistical methodology, data analysis algorithms and conventions remain the same as those in the original SAP.

2 Modification to the Multiplicity Adjustment

The hypotheses of interest for the planned labeling will be comprised of only the primary endpoint CMAI total score and the key secondary endpoint CGI-S score. The hypotheses pertaining to all other secondary endpoints will be excluded from the hierarchical testing procedure.

The testing will be conducted on two hypotheses of interest in the order specified below:

1) The primary efficacy endpoint, change from baseline to Week 12 in the CMAI total score

2) The key secondary efficacy endpoint, change from baseline to Week 12 in the CGI-S score

After reviewing the unblinded interim analysis results, DMC recommended that the trial be continued to the planned end. As a result, at the final analysis, the primary efficacy endpoint will be tested at a two-sided 3.5% significance level, and upon achieving significance on the primary endpoint, the key secondary endpoint will be tested at the same level. The hierarchical testing procedure controls the Type-I error rate for both the primary efficacy endpoint and the key secondary efficacy endpoint.

This page is a manifestation of an electronically captured signature

SIGNATURE PAGE

Document Name: P331-14-213_SAP v2.0 Addendum 08Sep2021

Document Number: [redacted]

Document Version: 2.0

Signed by	Meaning of Signature	Server Date (dd-MMM-yyyy hh:min) - UTC timezone
[redacted]	Biostatistics Approval	08-Sep-2021 15:56:42
[redacted]	Clinical Approval	08-Sep-2021 18:21:55

Otsuka Pharmaceutical Development & Commercialization, Inc.

Investigational Medicinal Product

Brexpiprazole (OPC-34712)

Addendum 2 to Statistical Analysis Plan Version 2.0

A Phase 3, 12-Week, Multicenter, Randomized, Double-blind, Placebo-controlled, 2-Arm, Fixed-dose Trial to Evaluate the Efficacy, Safety, and Tolerability of Brexpiprazole (OPC-34712) in the Treatment of Subjects With Agitation Associated With Dementia of the Alzheimer's Type

Protocol No. 331-14-213
IND No. 115,960
EudraCT No. 2017-003940-19

Version: Final

Date: 31 May 2022

CONFIDENTIAL - PROPRIETARY INFORMATION

Table of Contents

1 Introduction

This addendum to the statistical analysis plan (SAP) version 2.0 is to add the list of Extrapyramidal symptoms (EPS)-related adverse events (AEs) and document the changes in the Proposed List of Summary Tables after the approval of the original SAP (version 2.0, issued on 29 January 29, 2021). The addition and changes will have no impact on the familywise Type-I error rate, which is strongly controlled at a two-sided 5% level. The statistical methodology, data analysis algorithms and conventions remain the same as those in the original SAP version 2.0.

2 Addition of List of EPS-related AEs

EPS-related AEs are listed below and will be grouped into five categories.

1) Dystonic Events, which include blepharospasm, cervical spasm, clumsiness, dystonia, dystonic tremor, emprosthotonus, essential tremor, facial spasm, fumbling, gait inability, head titubation, intention tremor, Meige's syndrome, muscle contractions involuntary, muscle spasms, muscle spasticity, muscle tightness, muscle tone disorder, musculoskeletal stiffness, myotonia, nuchal rigidity, oculogyration, oesophageal spasm, opisthotonos, opisthotonus, oromandibular dystonia, oropharyngeal spasm, pharyngeal dystonia, pleurothotonus, risus sardonicus, spasmodic dysphonia, tongue spasm, torticollis, torticollis psychogenic, trismus, uvular spasm;

2) Parkinsonian Events, which include action tremor, akinesia, asterixis, bradykinesia, bradyphrenia, cogwheel rigidity, dysphonia, fine motor skill dysfunction, freezing phenomenon, gait disturbance, gait festinating, hypertonia, hypokinetic dysarthria, laryngeal tremor, masked facies, micrographia, mobility decreased, muscle rigidity, on and off phenomenon, parkinsonian crisis, parkinsonian gait, parkinsonian rest tremor, parkinsonism, parkinsonism hyperpyrexia syndrome, parkinsons disease, parkinson's disease, parkinson's disease psychosis, postural reflex impairment, postural tremor, propulsive gait, reduced facial expression, resting tremor, tremor, walking disability;

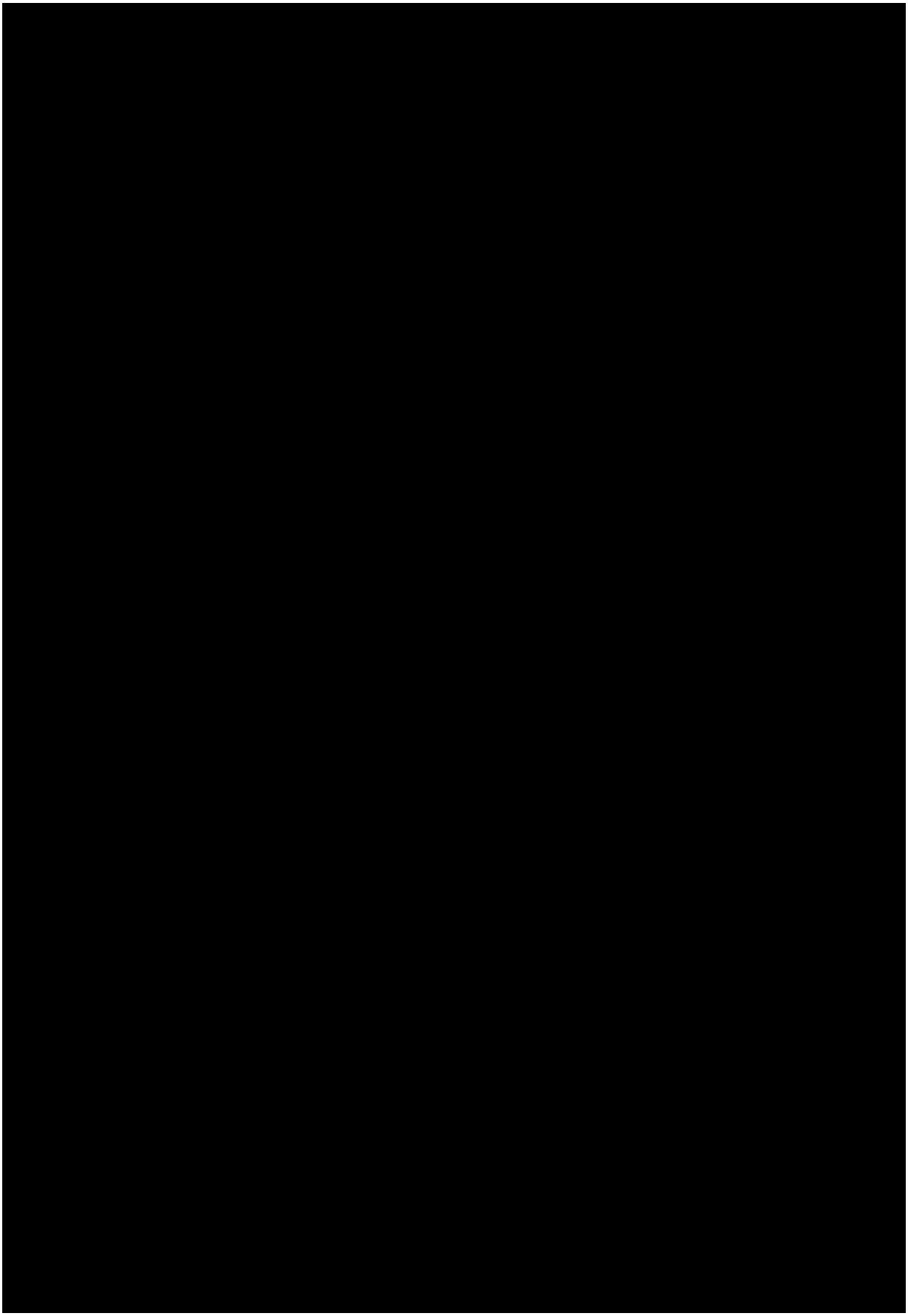
3) Akathisia Events, which include akathisia, extrapyramidal disorder, hyperkinesia, movement disorder, psychomotor hyperactivity, and restlessness;

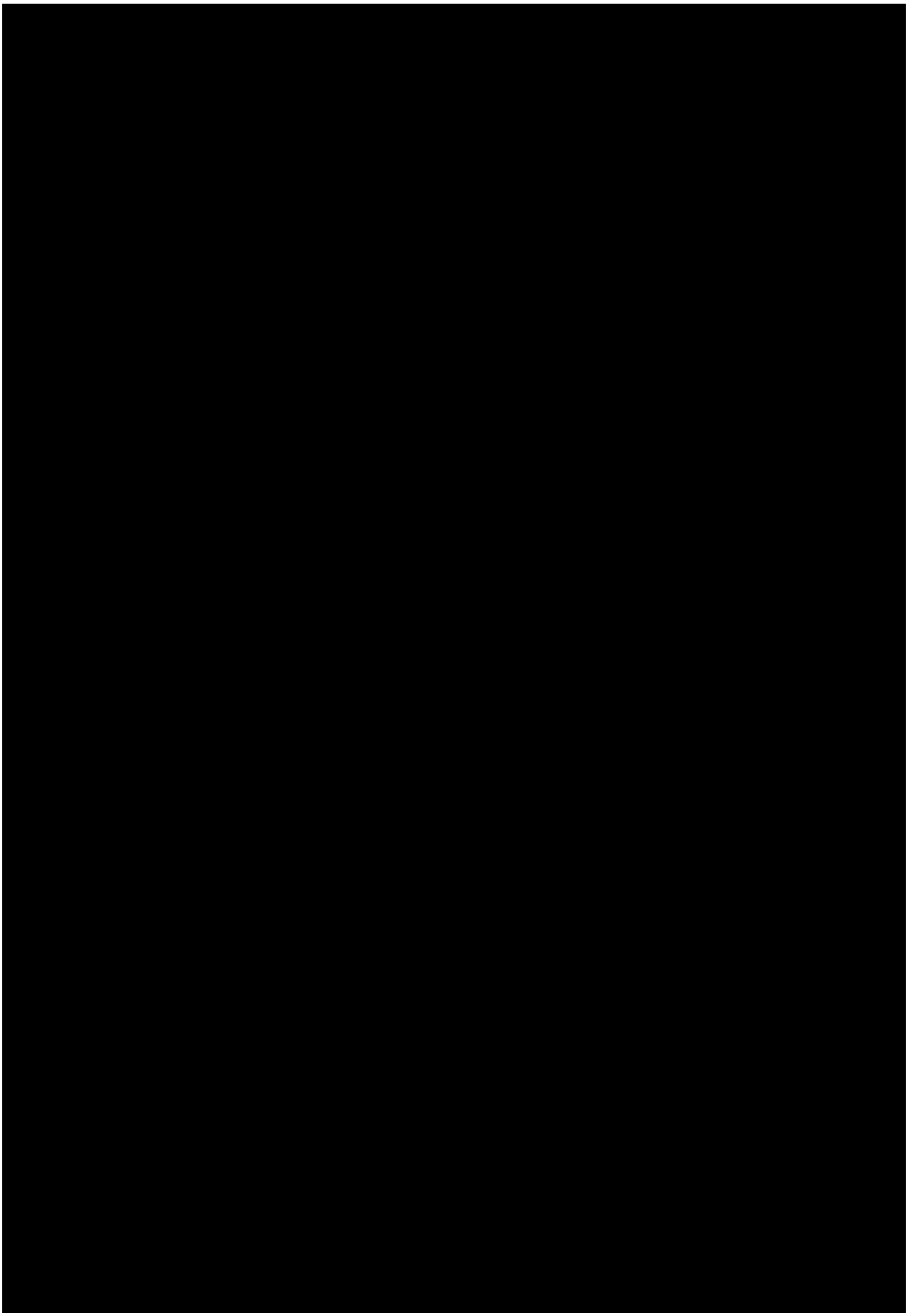
4) Dyskinetic Events, which include abnormal involuntary movement scale, athetosis, ballismus, buccoglossal syndrome, chorea, choreoathetosis, chronic tic disorder, complex tic, dopamine dysregulation syndrome, drooling, dyskinesia, dyskinesia oesophageal, grimacing, motor dysfunction, muscle twitching, nodding of head, oculogyric crisis,

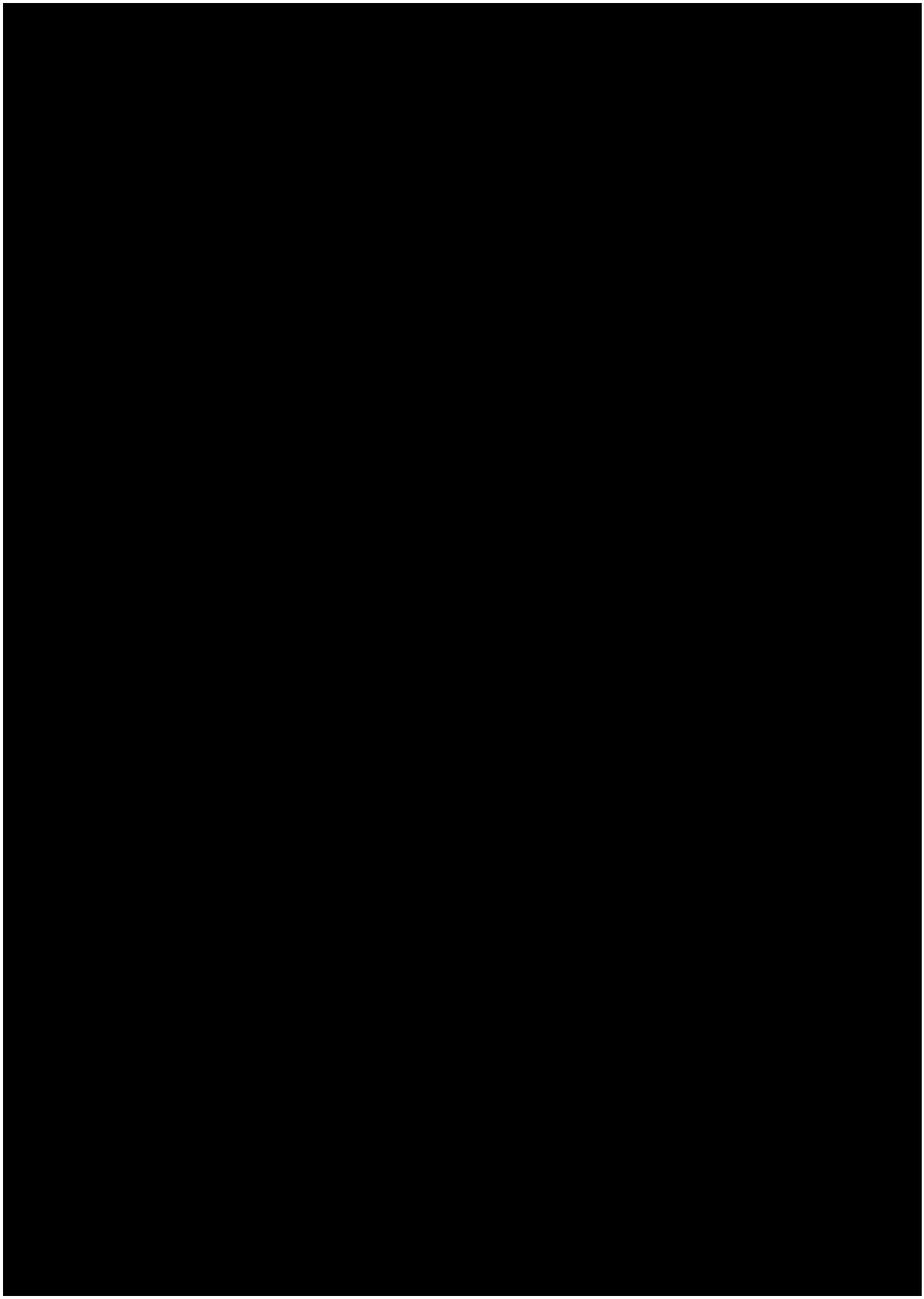
pharyngeal dyskinesia, protrusion tongue, provisional tic disorder, rabbit syndrome, respiratory dyskinesia, secondary tic, tardive dyskinesia, tic;

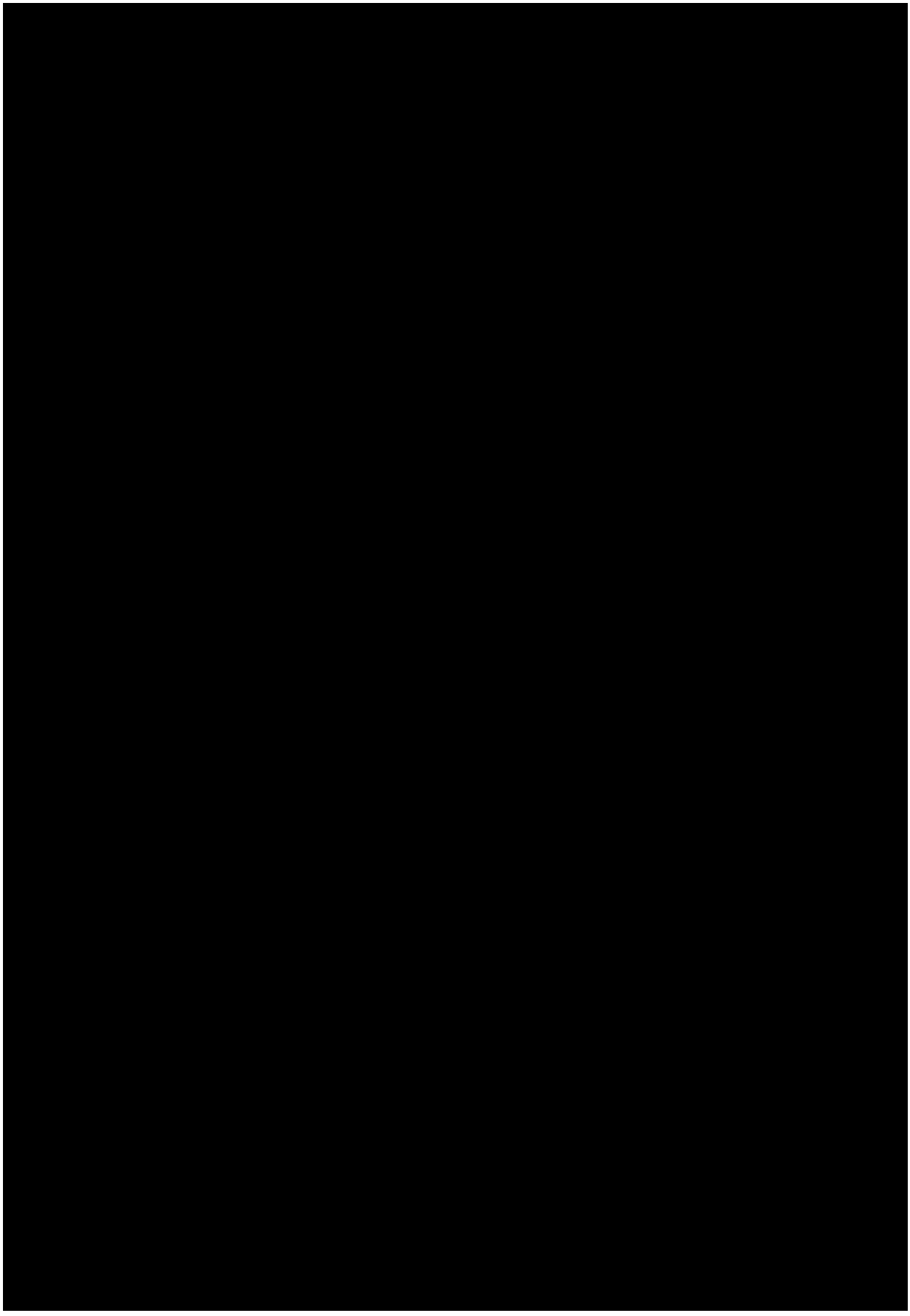
5) Residual Events, which include huntingtons disease and myoclonus.

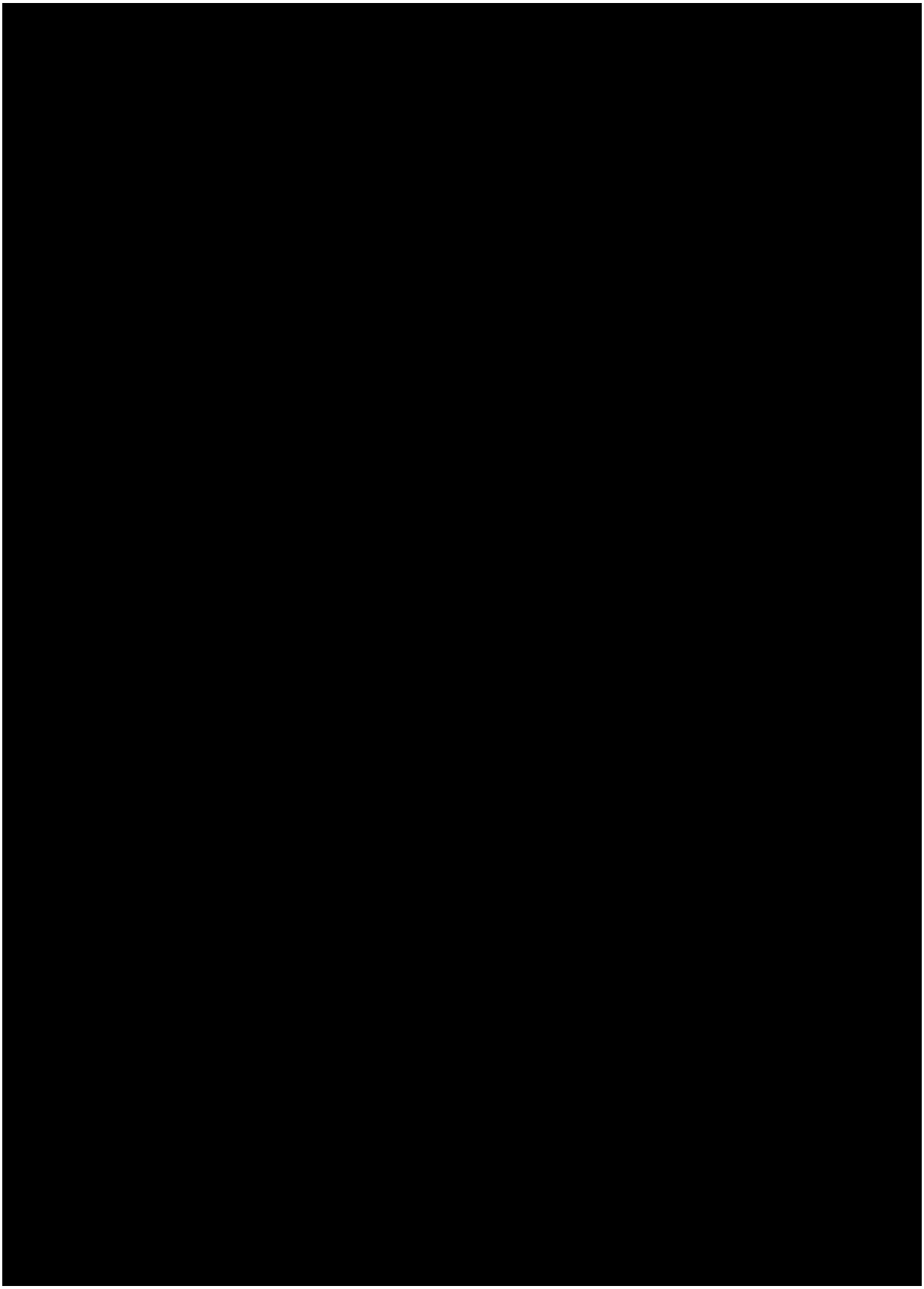
[redacted]

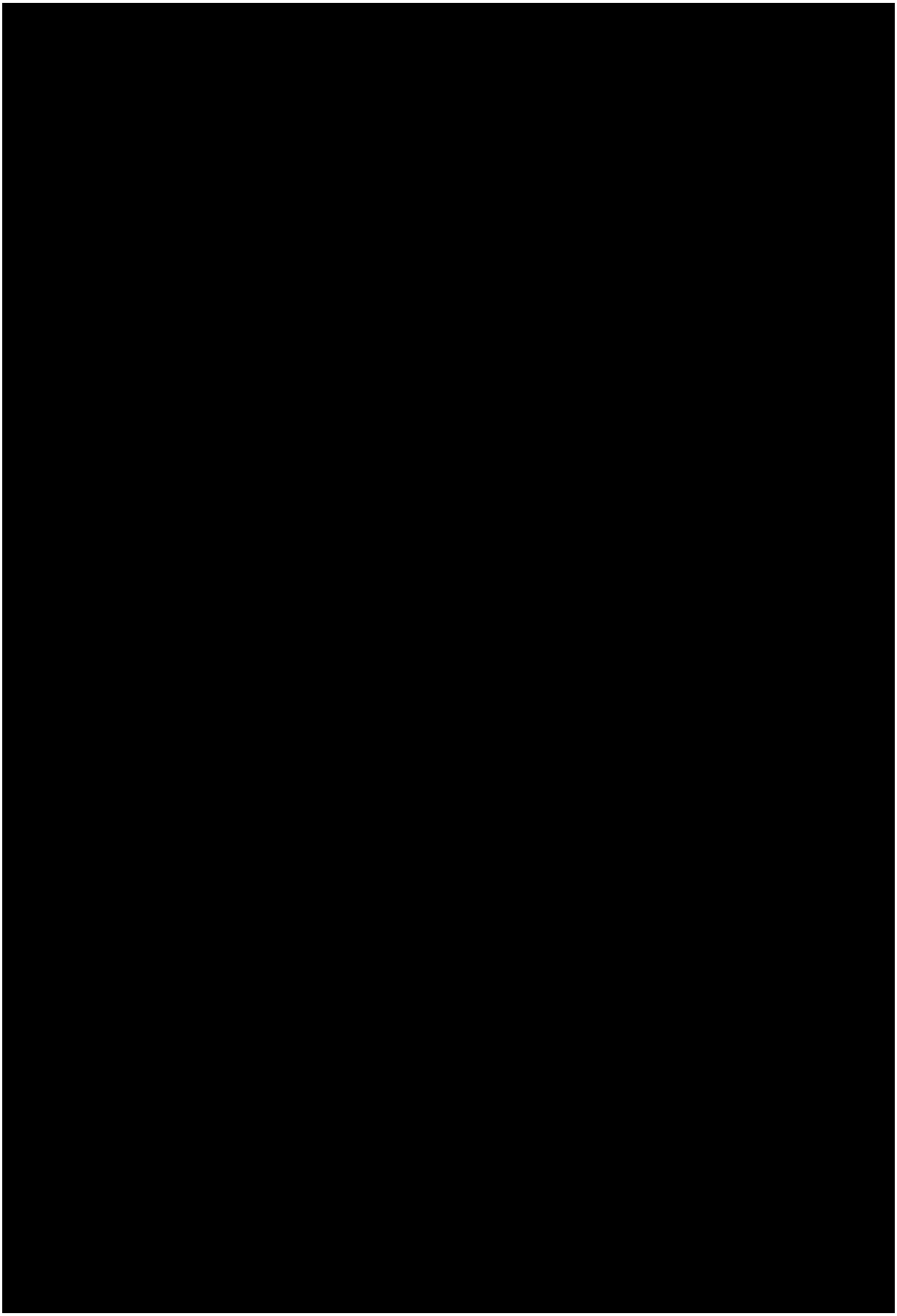


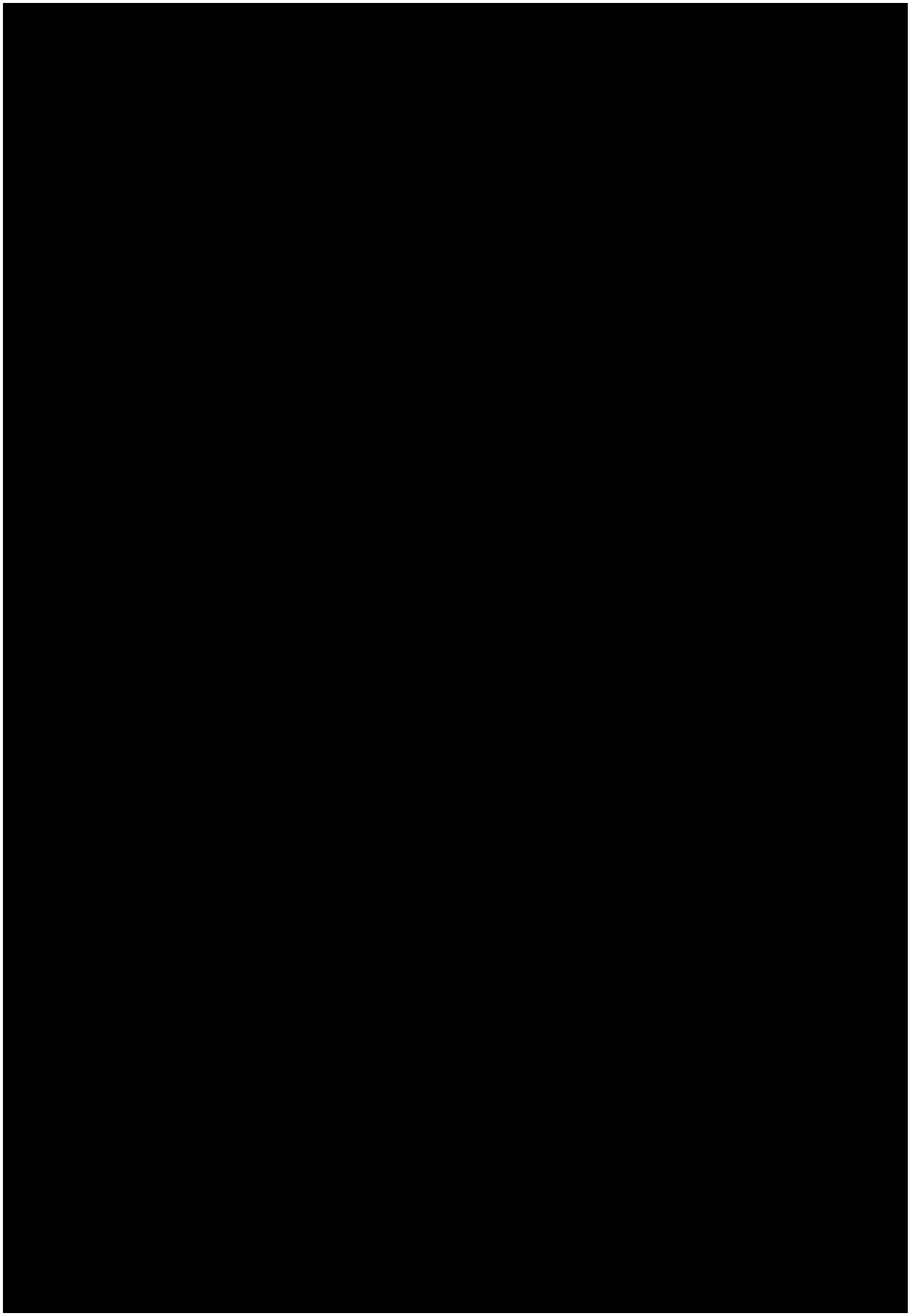












This page is a manifestation of an electronically captured signature

SIGNATURE PAGE

Document Name: P331-14-213_SAP v2.0 Addendum 2

Document Number: [redacted]

Document Version: 2.0

Signed by	Meaning of Signature	Server Date (dd-MMM-yyyy hh:min) - UTC timezone
[redacted]	Biostatistics Approval	31-May-2022 17:50:25
[redacted]	Clinical Approval	31-May-2022 15:16:32

Otsuka Pharmaceutical Development & Commercialization, Inc.

Investigational Medicinal Product

Brexpiprazole (OPC-34712)

INTERIM ANALYSIS PLAN

for

Protocol No. **331-14-213**
IND No. **115,960**
EudraCT No. 2017-003940-19

A Phase 3, 12-Week, Multicenter, Randomized, Double-blind, Placebo-controlled, 2-Arm, Fixed-dose Trial to Evaluate the Efficacy, Safety, and Tolerability of Brexpiprazole (OPC-34712) in the Treatment of Subjects With Agitation Associated With Dementia of the Alzheimer's Type

Version 4.0
Date: 10 September 2020

Protocol 331-14-213 Interim Analysis Plan

Table of Contents

List of Abbreviations and Definitions of Terms

Abbreviation	Definition
AAD	Agitation in Alzheimer's dementia
CGI-S	Clinical Global Impression - Severity of Illness
CMAI	Cohen-Mansfield Agitation Inventory
CP	Conditional power
DMC	Data Monitoring Committee
IA	Interim analysis
IAP	Interim analysis plan
IMP	Investigational medicinal product(s)
IWRS	Interactive web response system
LS	Least square
MMRM	Mixed-effect model repeated measure
OC	Observed case
OPDC	Otsuka Pharmaceutical Development & Commercialization, Inc.
OTC	Opportunity to complete
SAP	Statistical analysis plan
UN	Unstructured

Protocol 331-14-213 Interim Analysis Plan

1 Introduction

This interim analysis plan (IAP) documents the statistical methodology, data analysis algorithms, and conventions to be applied in the interim analysis (IA) and reporting of efficacy data of Trial 331-14-213.

Protocol 331-14-213 is a phase 3, 12-week, multicenter, randomized, double-blind, placebo-controlled, fixed-dose trial designed to assess the efficacy, safety, and tolerability of brexpiprazole compared with placebo. Subjects will be randomized in a 2:1 ratio to brexpiprazole or placebo. Within the brexpiprazole arm, subjects will be further randomized in a 1:2 ratio to 2 mg/day or 3 mg/day, to explore the efficacy, safety, and tolerability of 2 mg/day and 3 mg/day brexpiprazole versus placebo. The randomization will be stratified by site.

The primary objective is to compare the efficacy of brexpiprazole with placebo in subjects with agitation associated with dementia of the Alzheimer's type, as assessed by the change from baseline to week 12 in the Cohen-Mansfield Agitation Inventory (CMAI) total score.

The trial consists of a 2- to 42-day screening period, a 12-week double-blind treatment period, and a 30-day post-treatment safety follow-up period. In addition, for all subjects who terminate early from the trial, attempts will be made to collect data on mortality status by telephone contact with the subject's caregiver at Week 16.

2 Statistical Methods

2.1 Sample Size and Power Justification

A sample of 300 subjects was calculated based on the treatment effect of 6.5 points with a standard deviation of 16.5 in the change from baseline to Week 12 CMAI total score, to achieve 89% power at a 2-sided alpha level of 0.05. Taking into account an empirical dropout rate of 12-13% (based on two completed phase 3 AAD trials) and a lack of compliance at one identified clinical site, an increment of at least 30 additional subjects is needed to maintain the target power. The resulting sample size is about 330 total subjects to be randomized in a 2:1 ratio to brexpiprazole and placebo. Within the brexpiprazole arm, subjects will be further randomized in a 1:2 ratio to 2 mg/day and 3 mg/day, to explore the efficacy, safety, and tolerability of 2 mg/day and 3 mg/day brexpiprazole versus placebo.

Protocol 331-14-213 Interim Analysis Plan

One interim look will be conducted when approximately the first 255 subjects have had an opportunity to complete (OTC) the 12-week trial. A Bonferroni boundary will be used for the two-stage group sequential analysis. The alpha allocated to the IA is 0.015 and the alpha left for the final analysis is 0.035, both two-sided. The power of the group-sequential test is about 87%.

The statistical assumption of the treatment effect of 6.5 points separation (for brexpiprazole versus placebo) is the same as that used for the two completed phase 3 AAD Trials 331-12-283 and 331-12-284.

2.2 Interim Analysis

A DMC will provide oversight for safety monitoring and reviewing the IA data.

An unblinded interim analysis of efficacy data is planned during the course of the trial and will be performed by an independent Data Monitoring Committee (DMC) when approximately the first 255 subjects have had an opportunity to complete the 12-week trial (i.e., either completed or discontinued from the trial before Week 12). Using a Bonferroni critical boundary for the rejection of the null hypothesis, a significance level of 0.015 (2-tailed) will be allocated to the interim analysis, and the significance level for the final analysis will be conservatively set at 0.035 (2-tailed).

Should the interim analysis be performed for a different number of subjects, the alpha allocation will remain unchanged for the interim and the final analysis.

For ethical concerns as well as to avoid wasting resources, a non-binding futility boundary for the IA will be implemented as follows. If a standardized effect size of $\leq$ 0.10 is observed at the IA, the DMC should recommend terminating the trial for futility.

The DMC will make a recommendation about stopping the trial at the IA or continuing the trial as planned based on the efficacy review. The results of the IA and individual subject data will remain blinded to the sponsor during the course of the trial and only DMC's recommendation to stop the trial for efficacy or to stop the trial for futility or to continue the trial to the planned maximum sample size will be communicated to the sponsors. The details of the DMC structure and its roles and responsibilities will be documented in the DMC charter.

The primary efficacy endpoint is the change from the baseline (Day 0 visit) to the end of the double-blind treatment period (the Week 12 visit) in the CMAI total score. The primary analysis will be performed on the portion of efficacy sample consisting of all

randomized subjects included in the IA who took at least one dose of IMP in the double-blind treatment period and had both a baseline and at least one post-randomization CMAI total score during the double-blind treatment period. The primary efficacy analysis will be performed by fitting a mixed-effect model repeated measure (MMRM) analysis with an unstructured (UN) variance covariance structure in which the change from the baseline in CMAI total score (Weeks 2, 4, 6, 8, 10, 12) will be the dependent variable based on the observed data. The model will include fixed class-effect terms for IMP (brexpiprazole and placebo), trial center, visit week, and an interaction term of IMP by visit week and include the interaction term of baseline values of CMAI Total Score by visit week as covariates. The primary comparison between the brexpiprazole (2-3 mg/day) and the placebo arm at Week 12 will be estimated as the difference between Least Squares (LS) means from the interaction term of IMP by visit week utilizing the computing software SAS procedure PROC MIXED.

Based on the IA result, the DMC will make one recommendation out of the three as follows:

1) Early stop for efficacy
2) Early stop for futility
3) Continue to the final analysis

IA result	DMC Recommendation	Outcome
Two-sided *p*-value ≤ 0.015	Stop the trial for superiority in efficacy	The trial will be terminated at IA
Observed effect size ≤ 0.10	Stop the trial for futility	The trial may be terminated at IA subject to Sponsors' final decision
Otherwise (None of the above)	Continue the trial to the final analysis	The trial will proceed to the planned maximum sample size of approximately N=330

The sponsors will remain blinded to any modification of the primary model due to non-convergence and should not be informed of any additional analysis or modification requested by the DMC and performed by the unblinded statistician.

3 Data Analysis

All interim efficacy and monitoring will be performed by the DMC. Reference is made to the DMC Charter for details of constituents and roles and responsibilities of DMC.

Protocol 331-14-213 Interim Analysis Plan

The DMC (Unblinded) Statistician will work closely with the DMC and will assist in the retrieval and analysis of efficacy data. The DMC will follow rules delineated in the IAP and the DMC Charter in disposing its responsibilities.

3.1 Planned Interim Analyses of Unblinded Efficacy data

A set of pre-specified efficacy tables (in Appendix) will be produced at the planned IA following statistical methods described in the Trial 331-14-213 Statistical Analysis Plan (SAP). The OPDC Biometrics Group will provide blinded SAS datasets and SAS programming codes to the DMC Statistician. The unblinded randomization codes will be provided to DMC Statistician directly by interactive web response system (IWRS) vendor. The DMC Statistician will produce all prespecified summary tables using the SAS codes and analysis datasets provided by the OPDC Biometrics Group.

4 Confidentiality of Results

Except for the situation when the DMC recommends stopping the trial because of compelling efficacy or safety reasons, all deliberations, results and findings of the DMC will remain confidential within their jurisdictions, and the sponsor and investigators will remain blinded of any or all findings of unblinded analysis of efficacy and safety data.

Protocol 331-14-213 Interim Analysis Plan

This page is a manifestation of an electronically captured signature

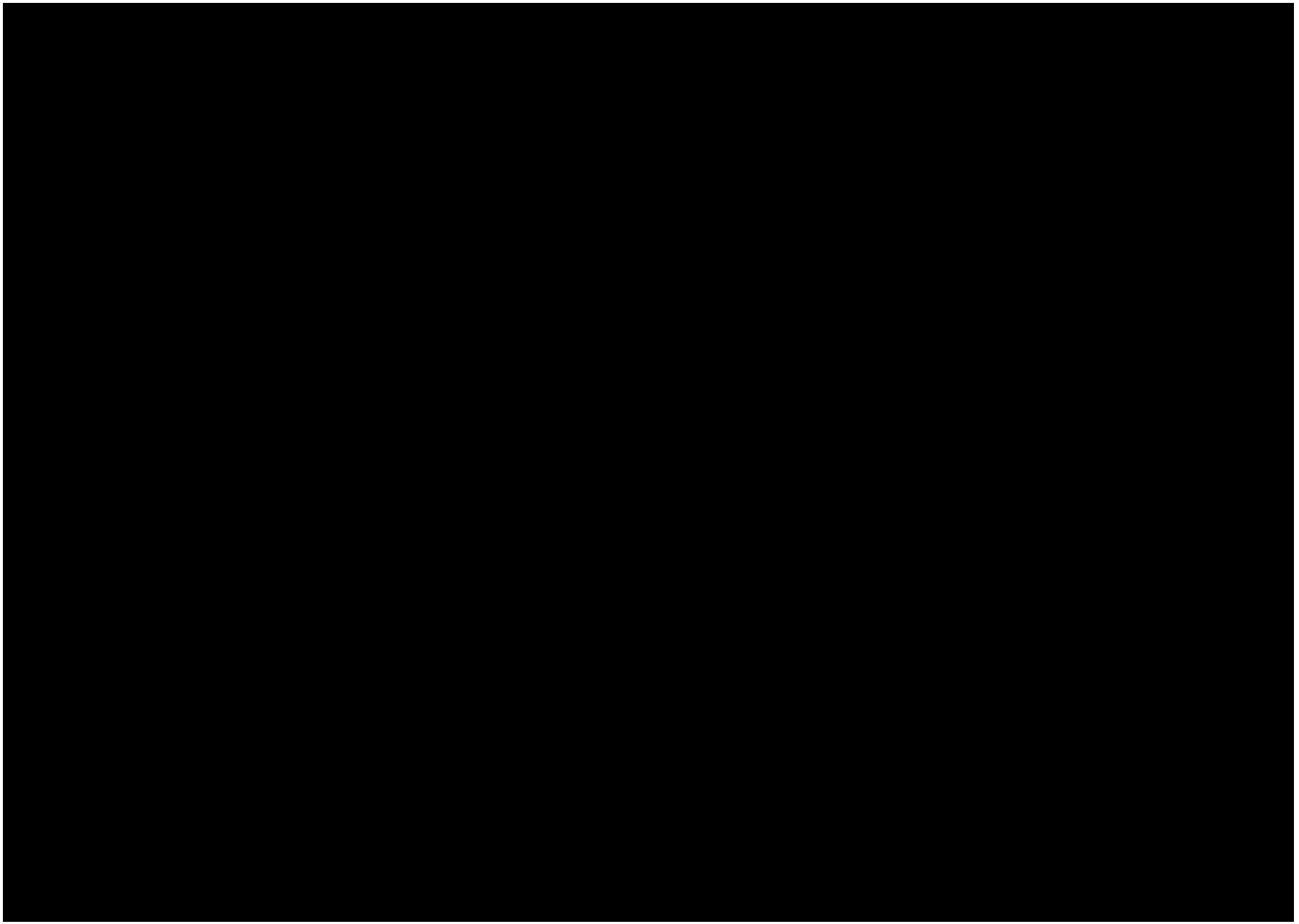
SIGNATURE PAGE

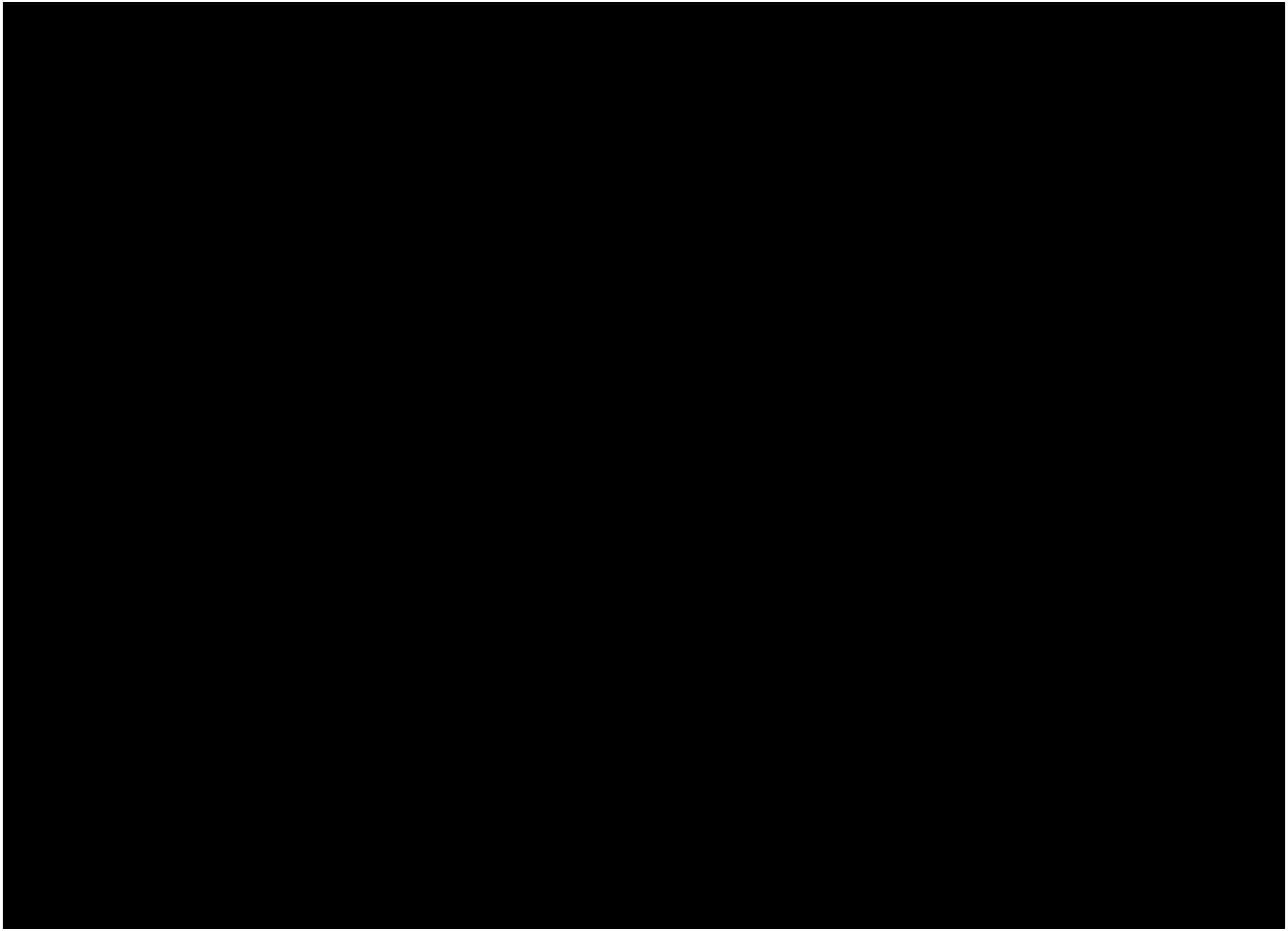
Document Name: P33114213_InterimAnalysisPlan V4.0

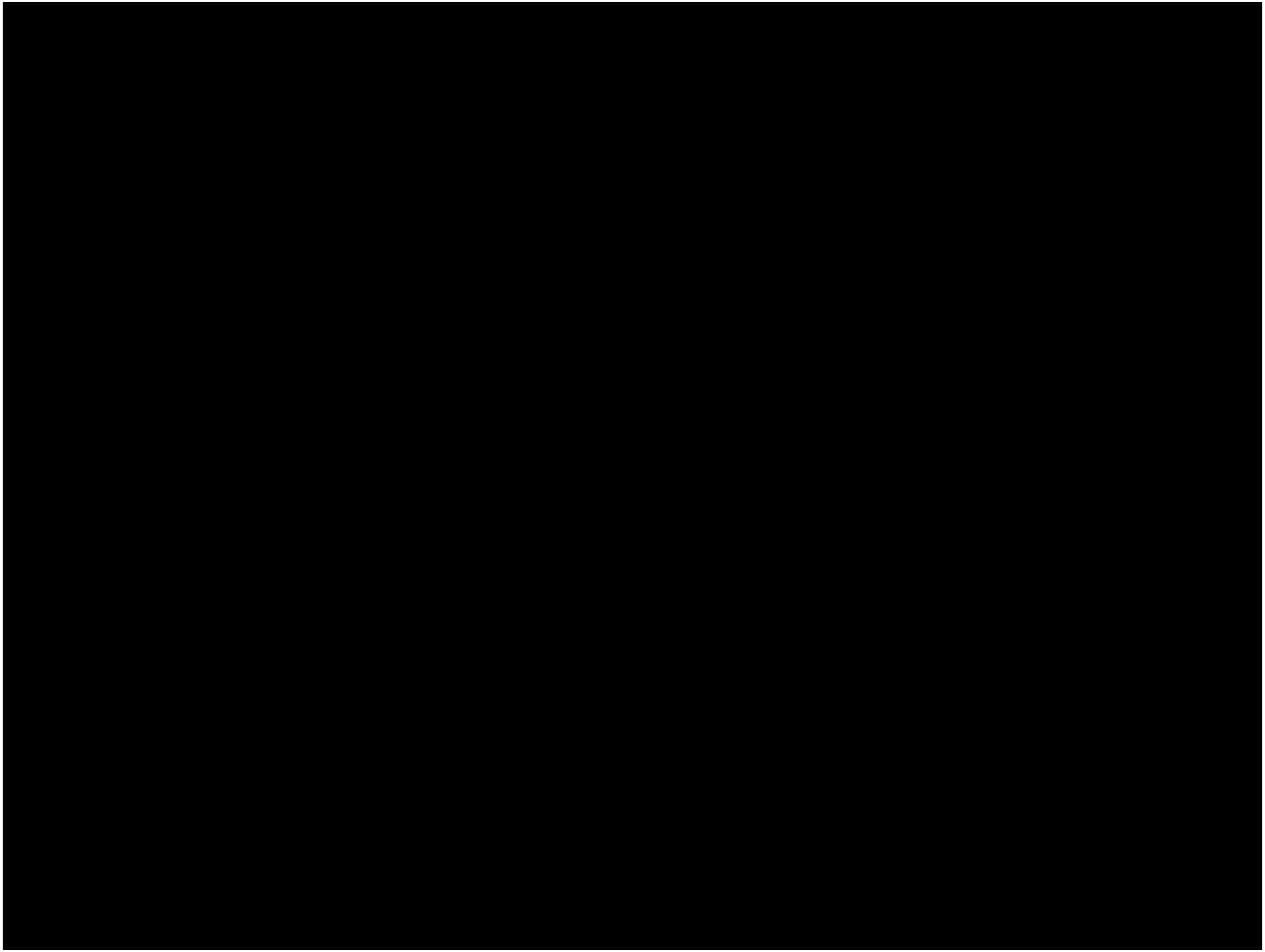
Document Number: [redacted]

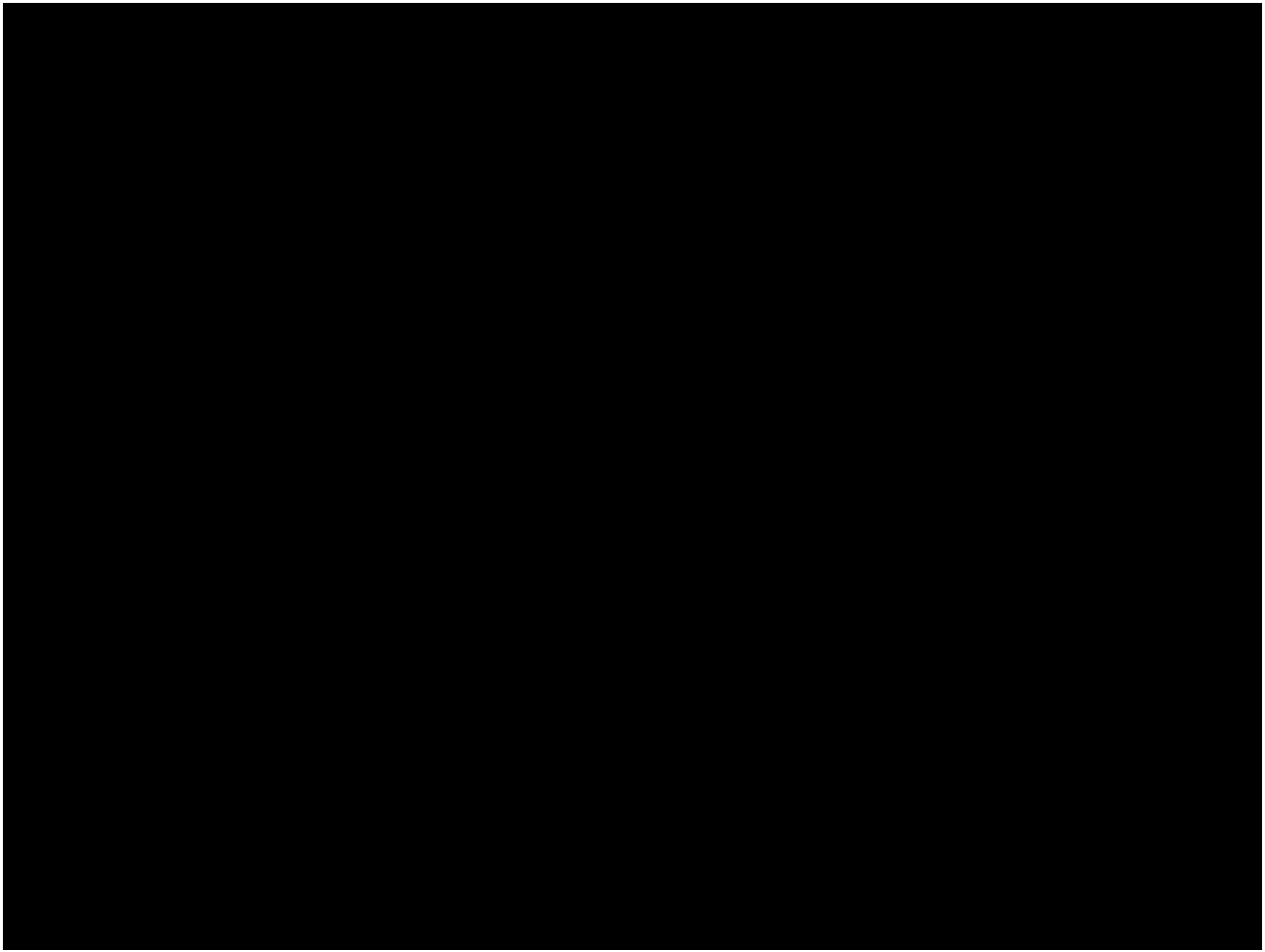
Document Version: 2.0

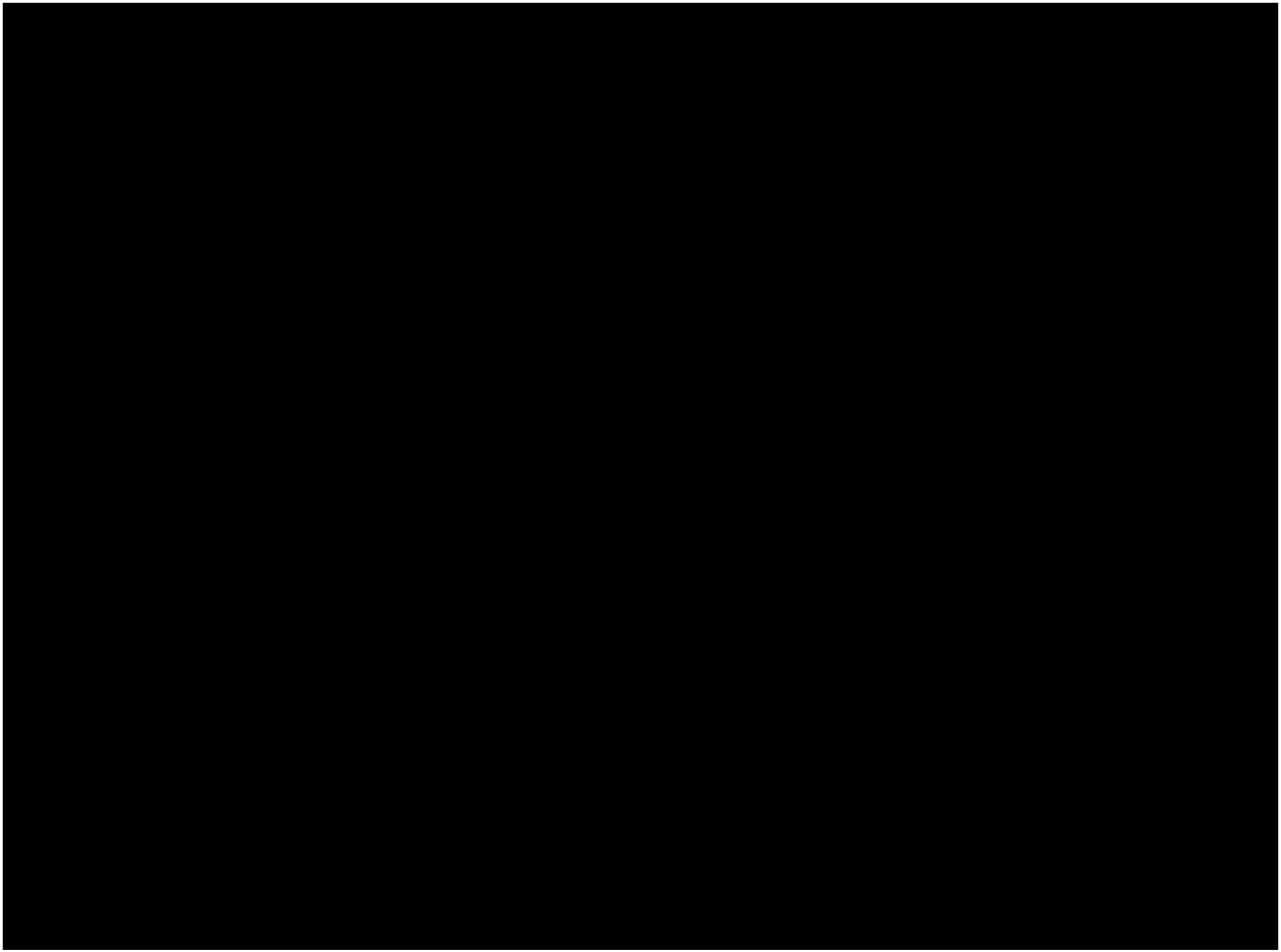
Signed by	Meaning of Signature	Server Date (dd-MMM-yyyy hh:min) - UTC timezone
[redacted]	Biostatistics Approval	10-Sep-2020 21:31:21
[redacted]	Clinical Approval	10-Sep-2020 20:19:01

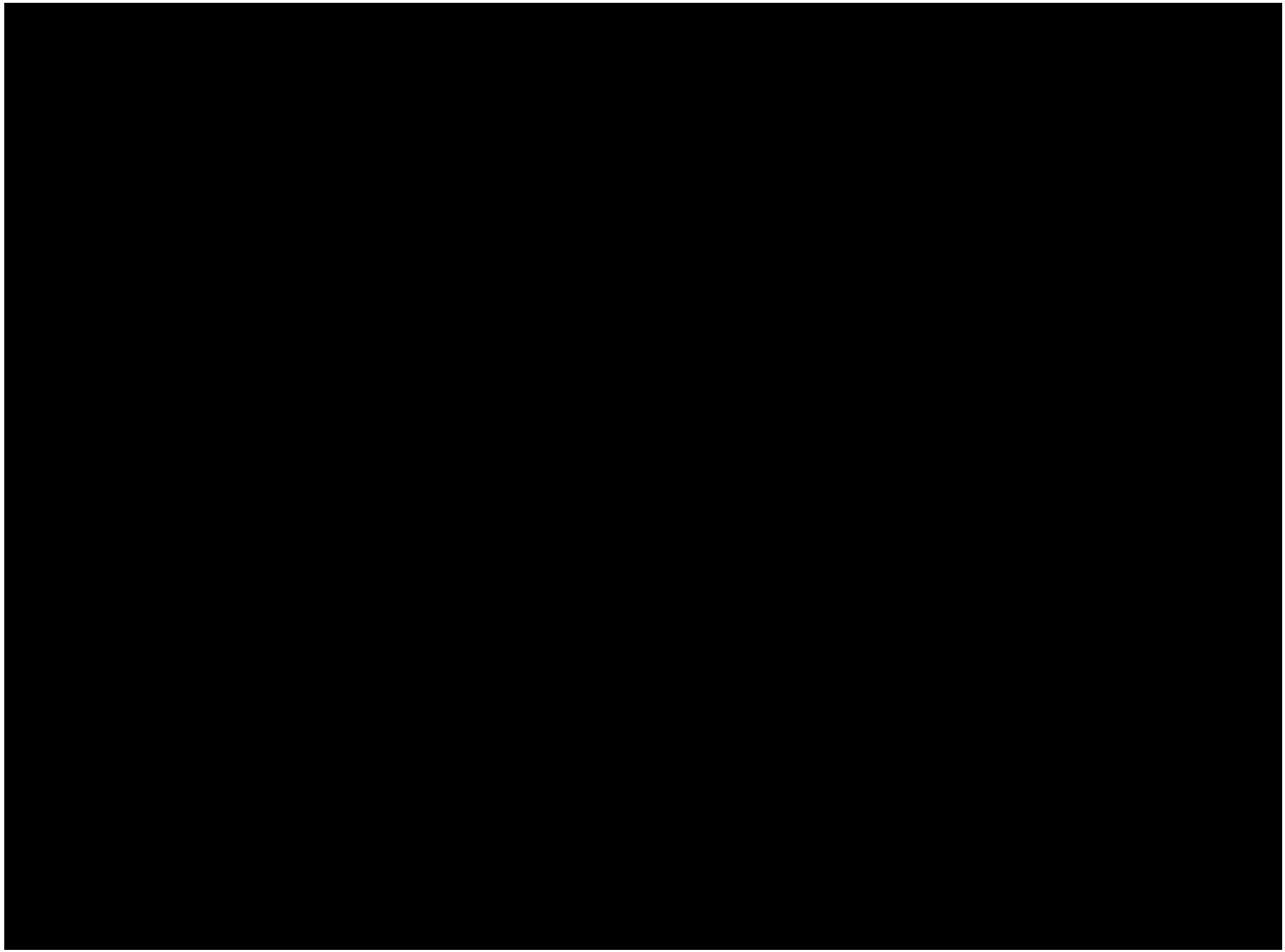


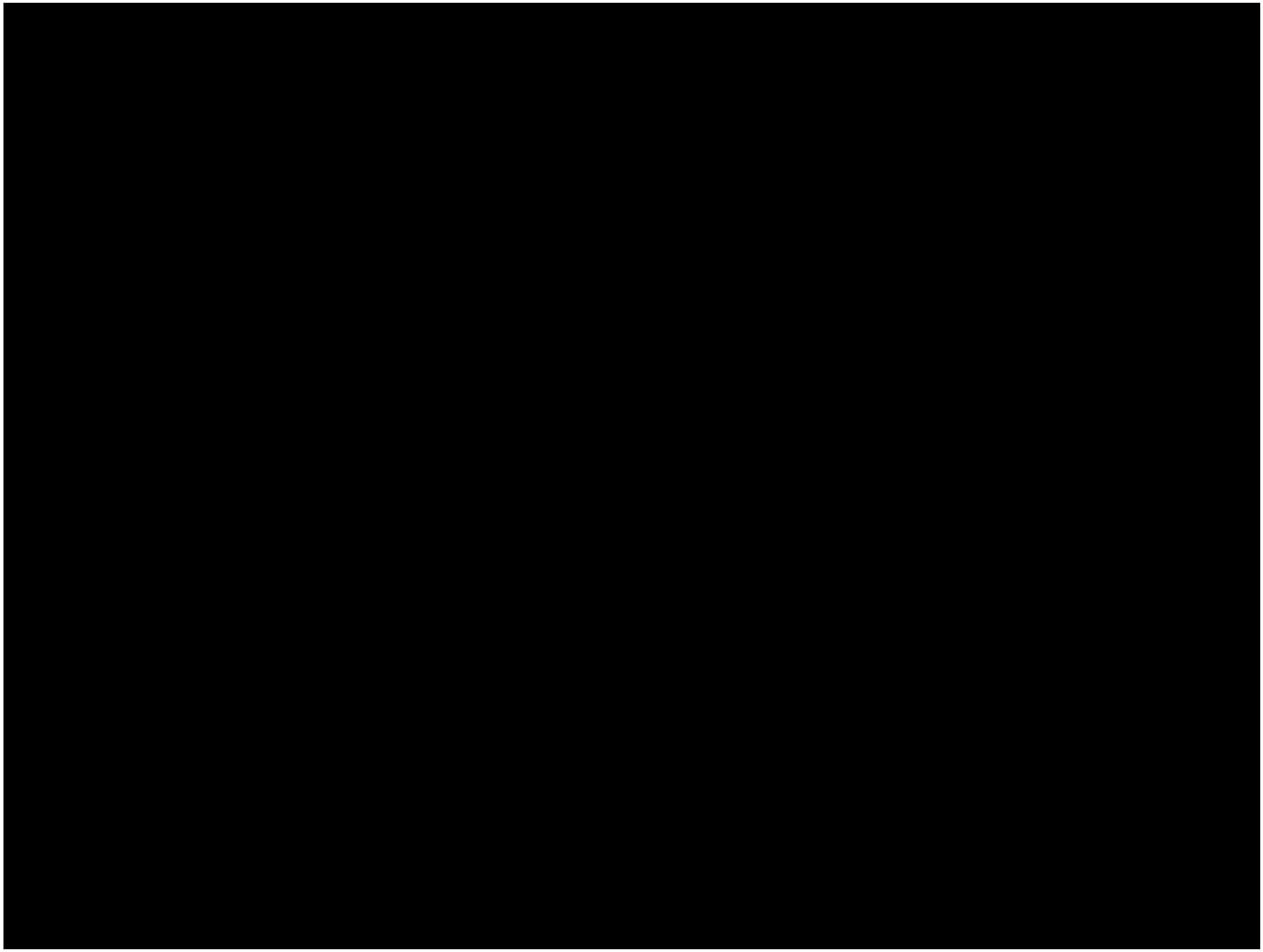


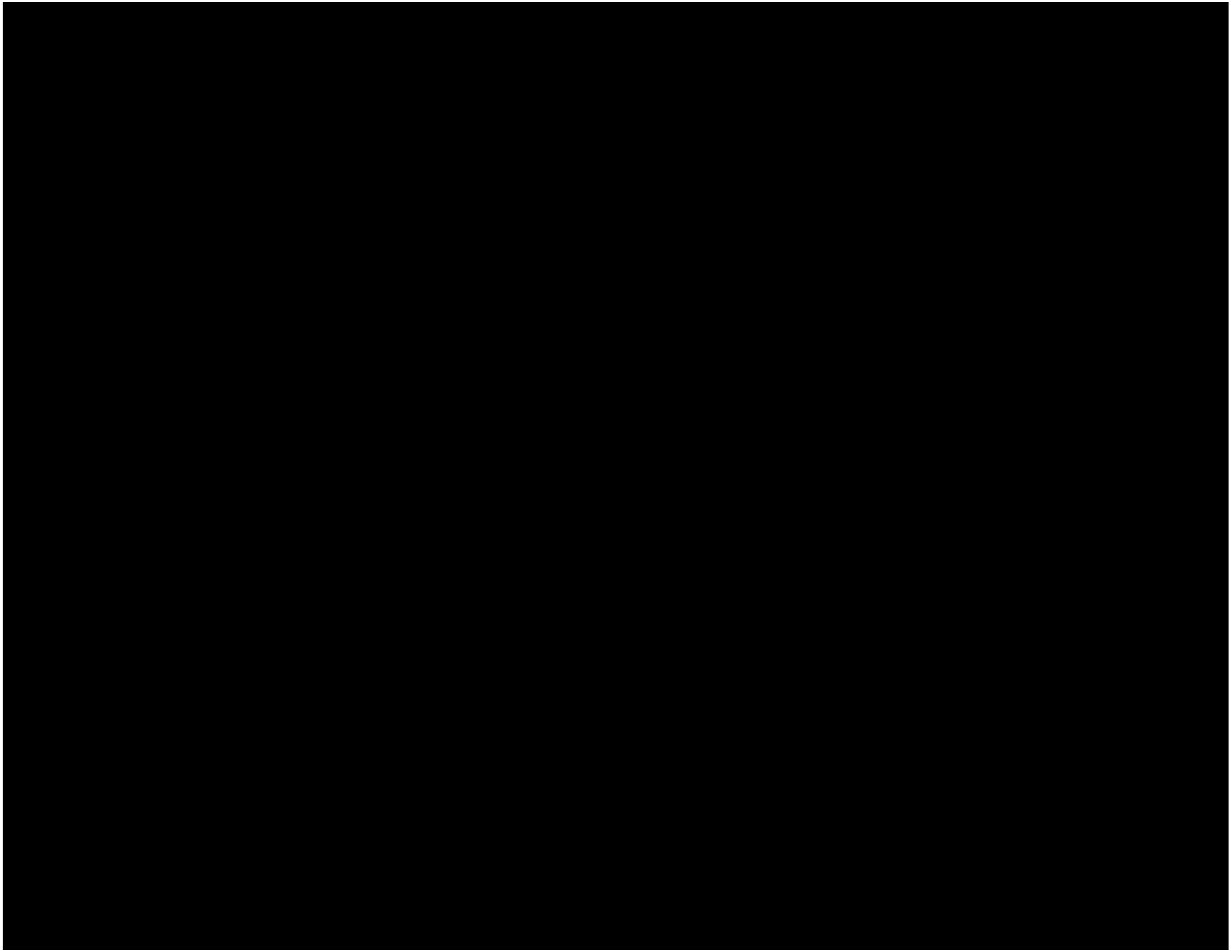


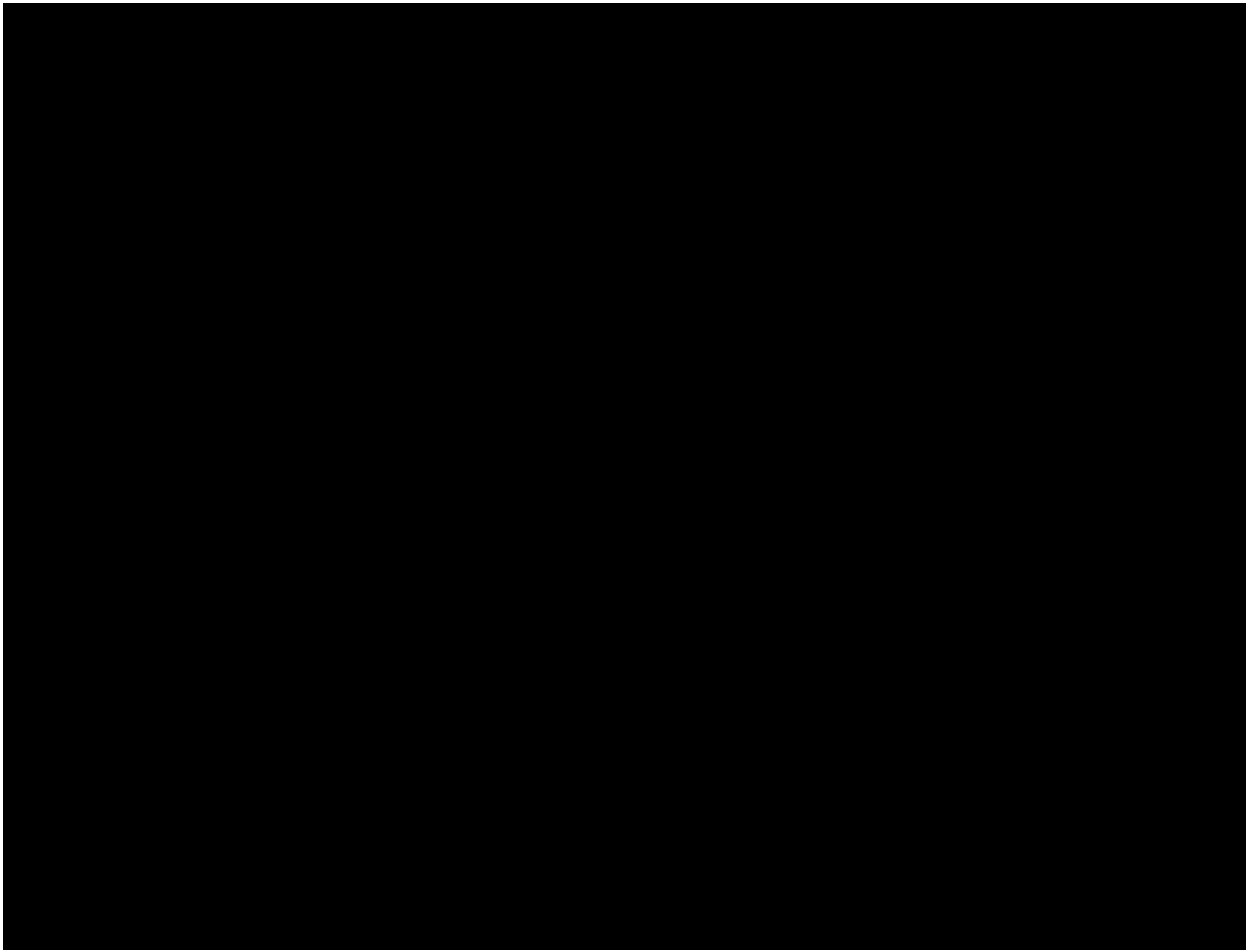


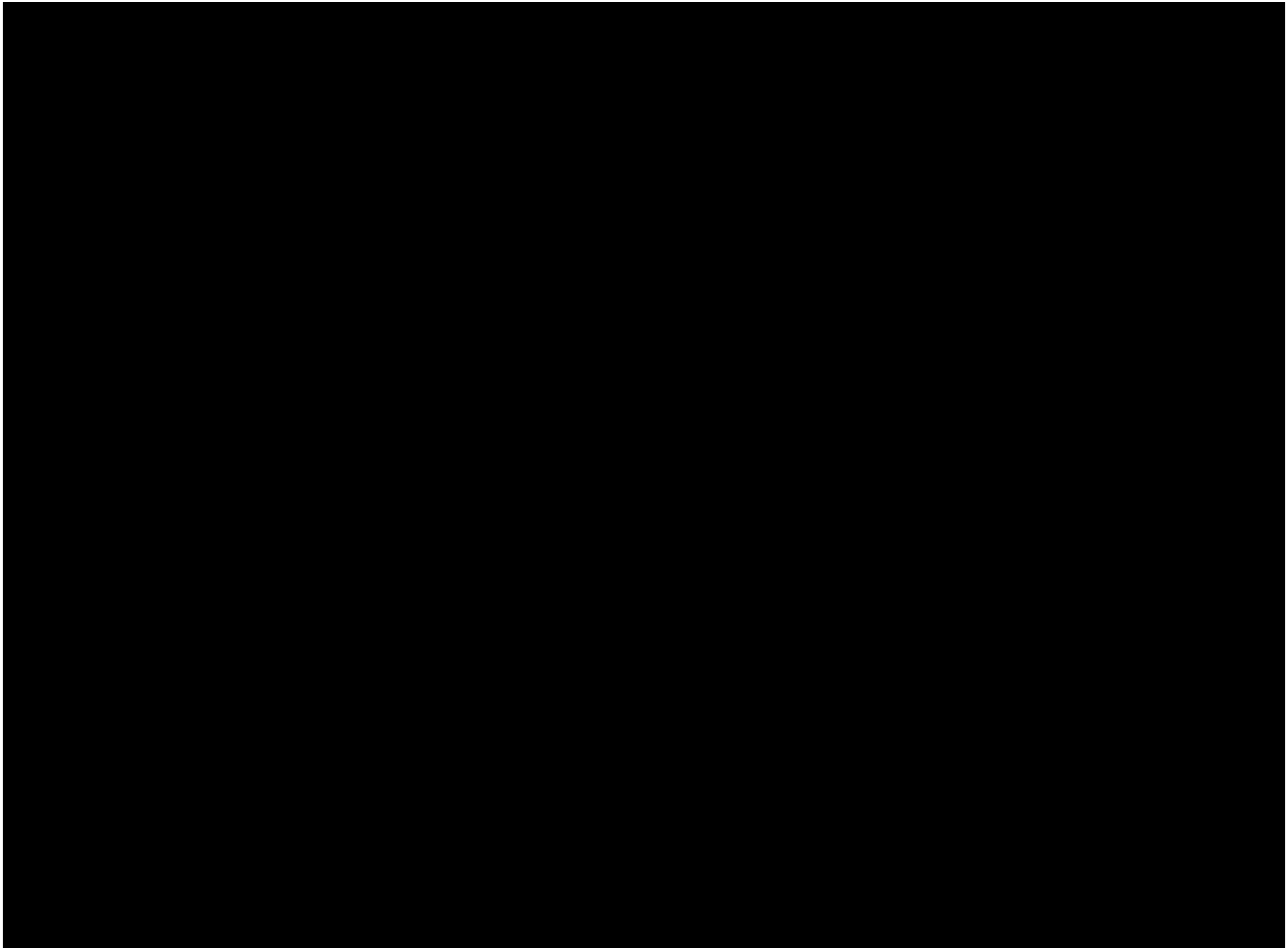


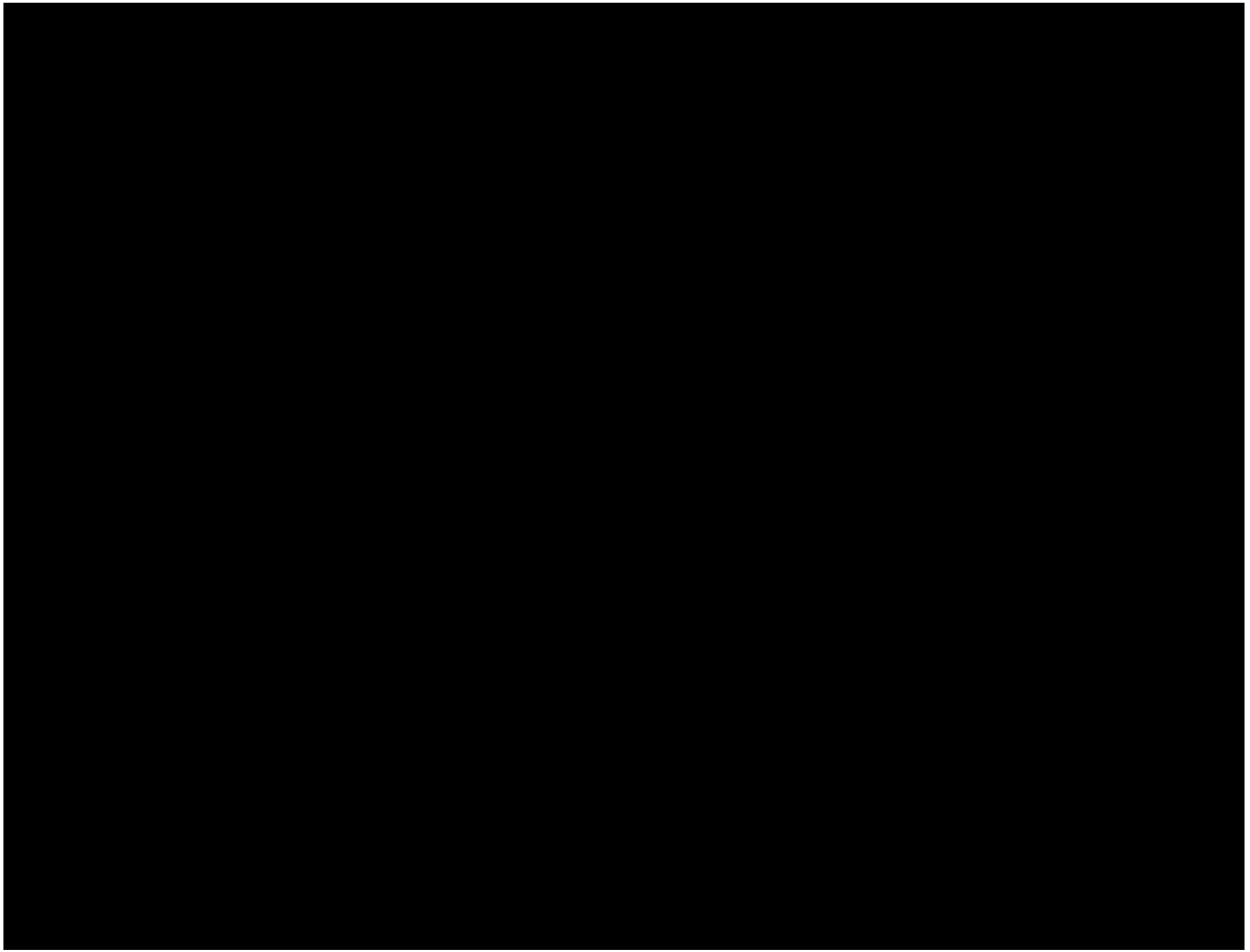


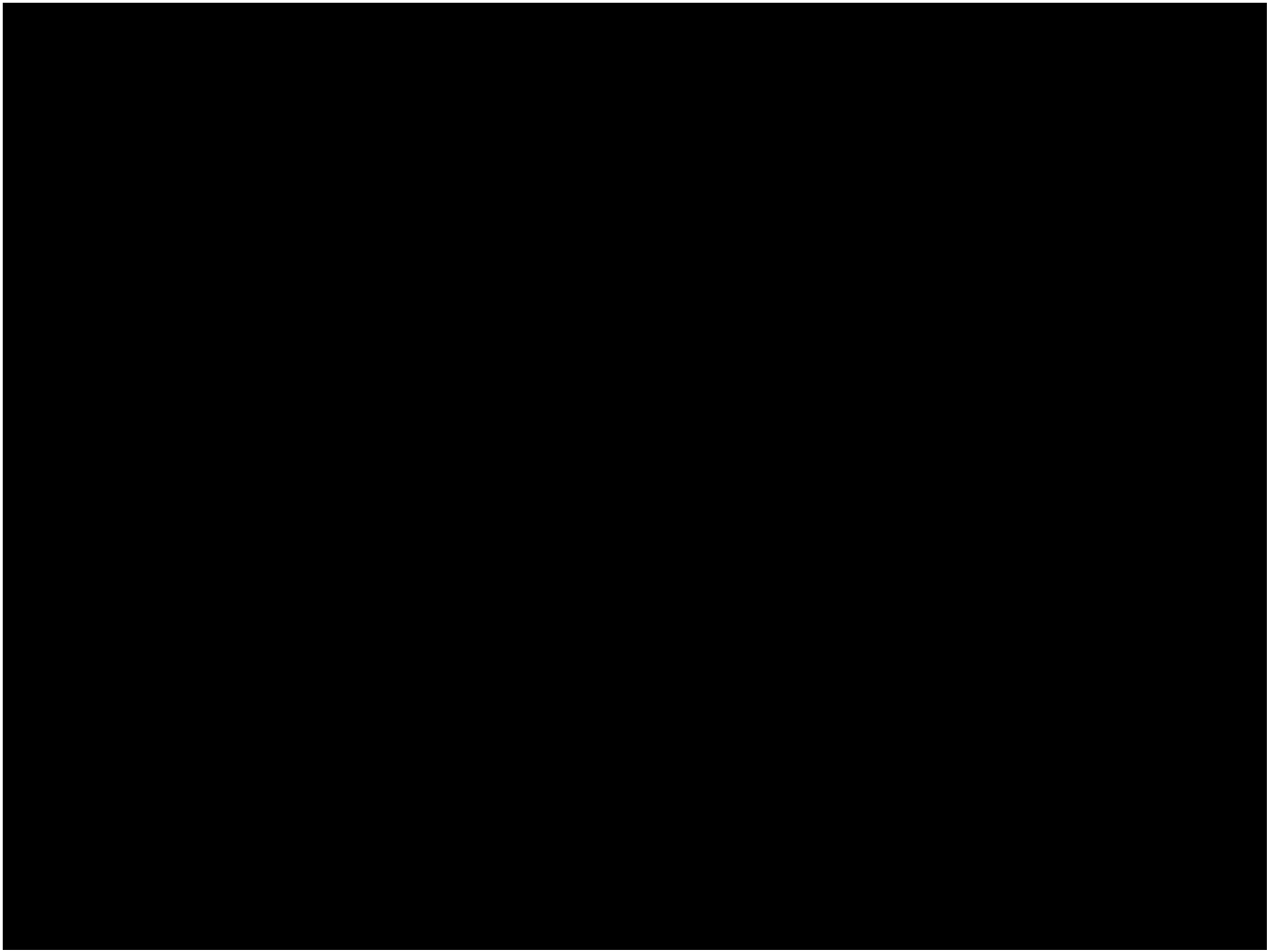


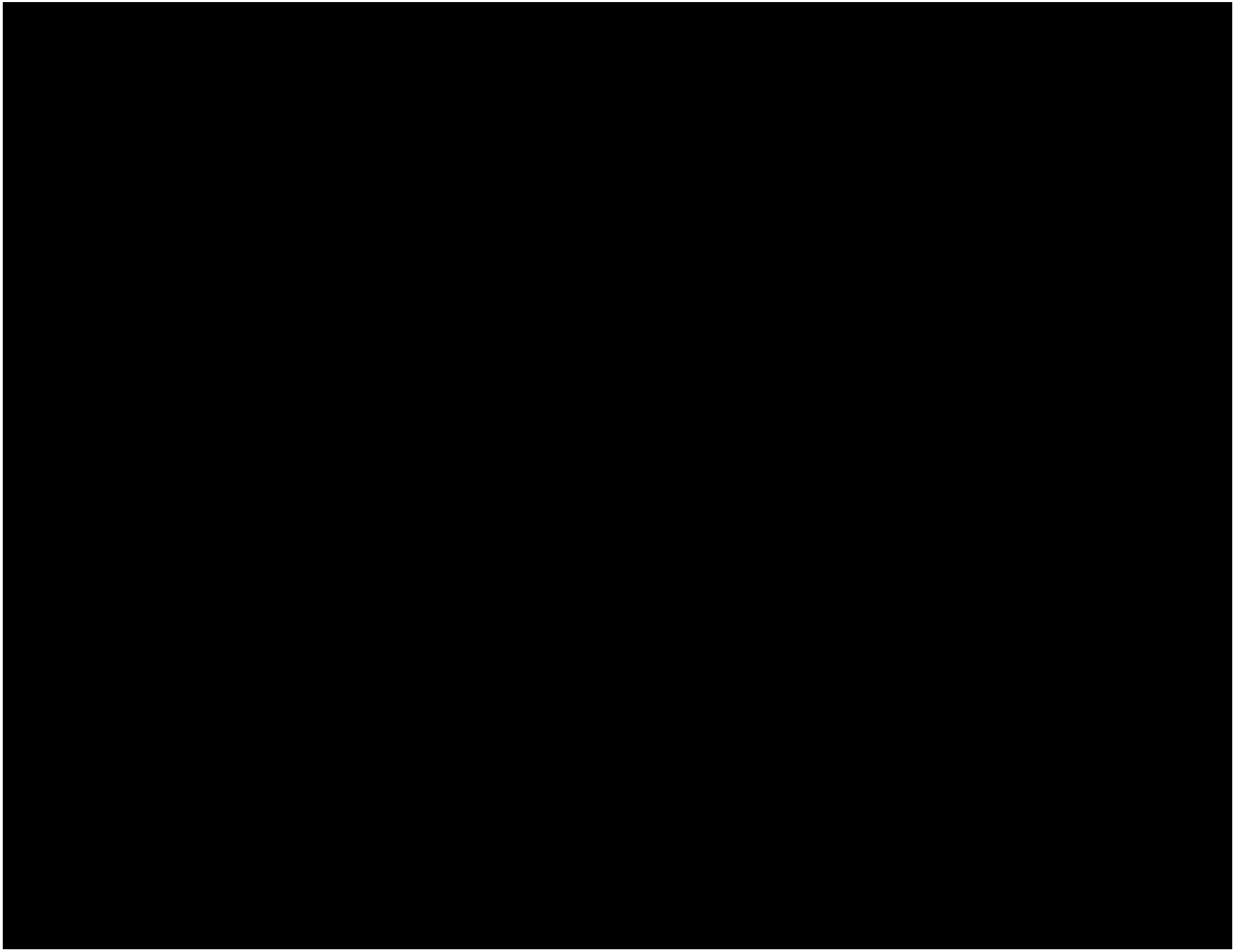


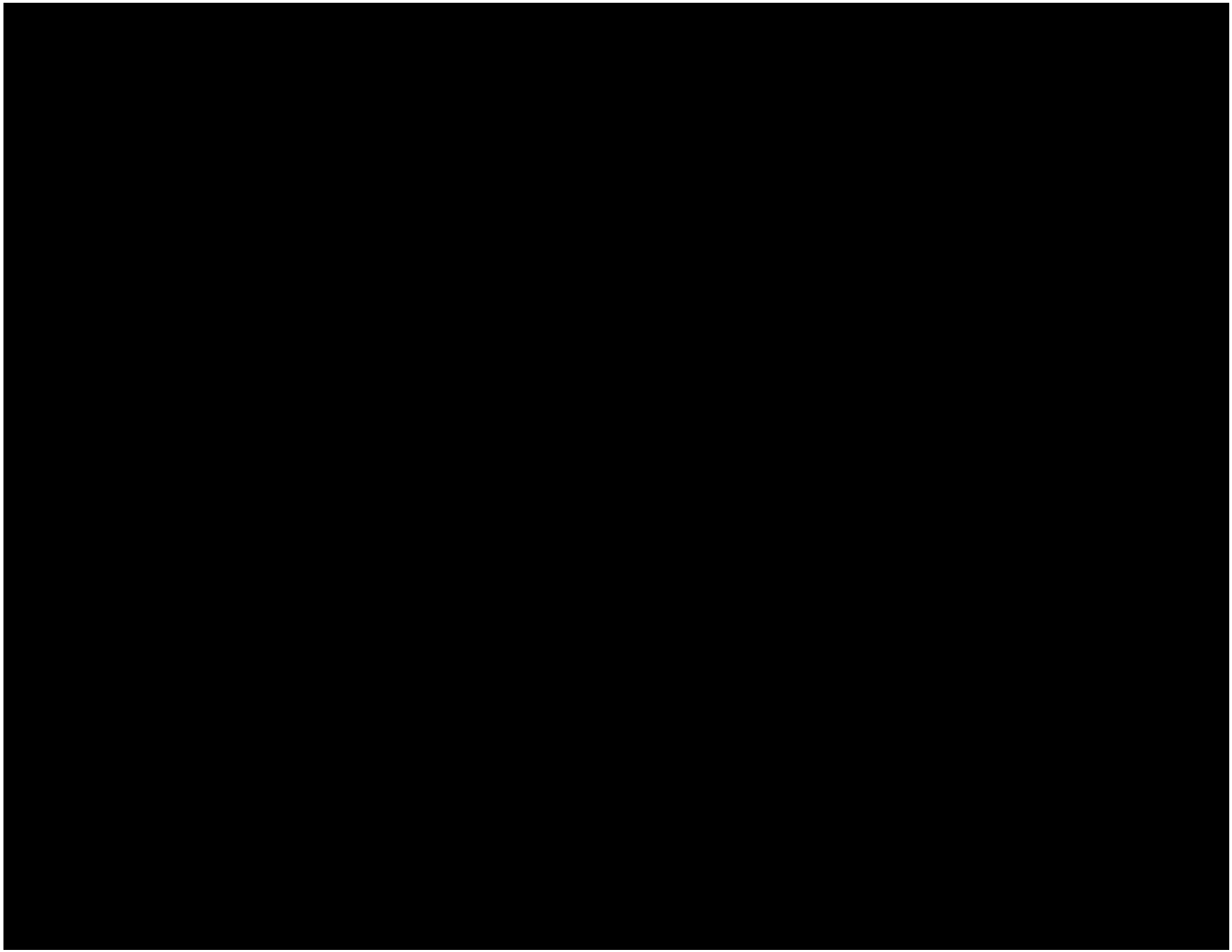


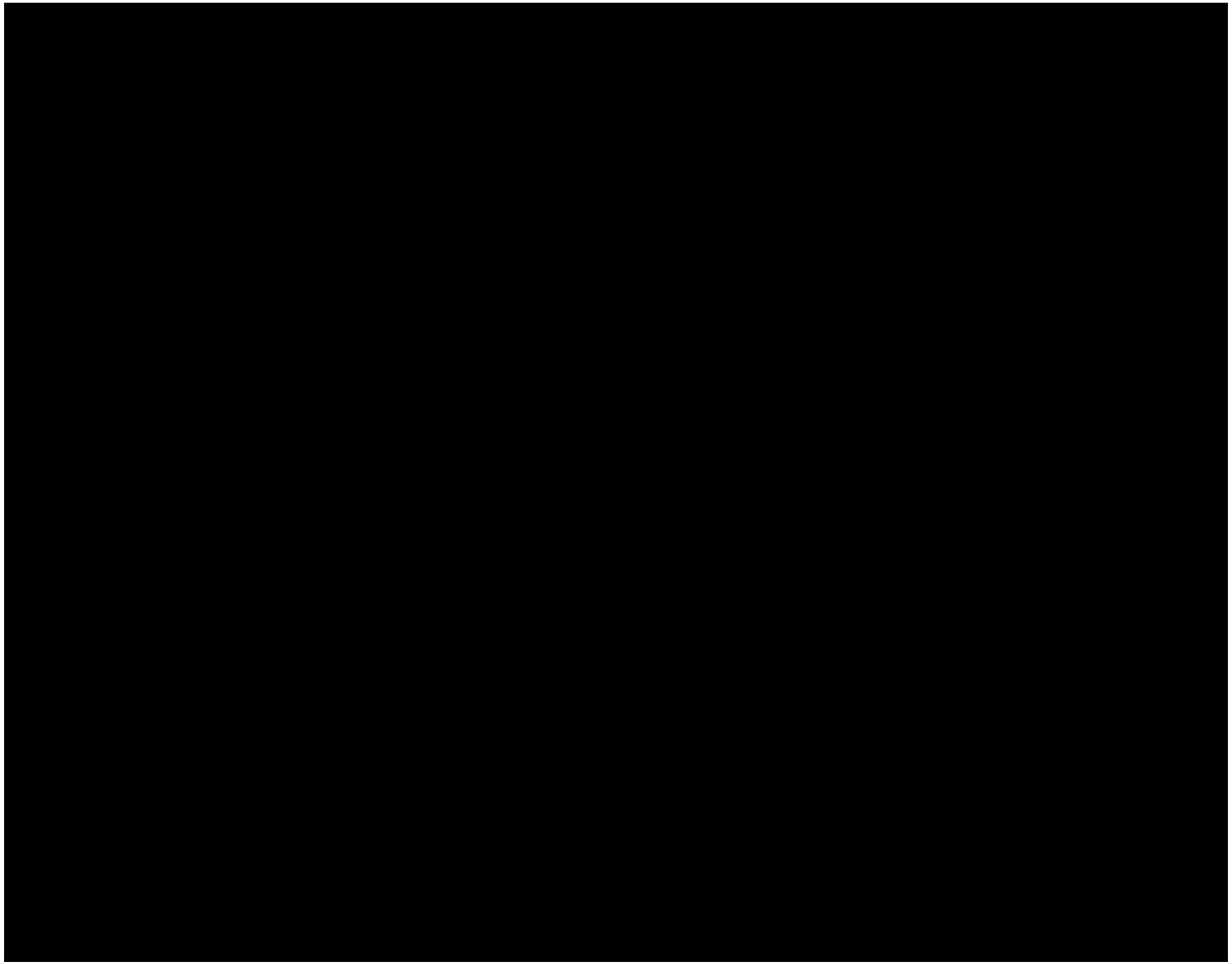


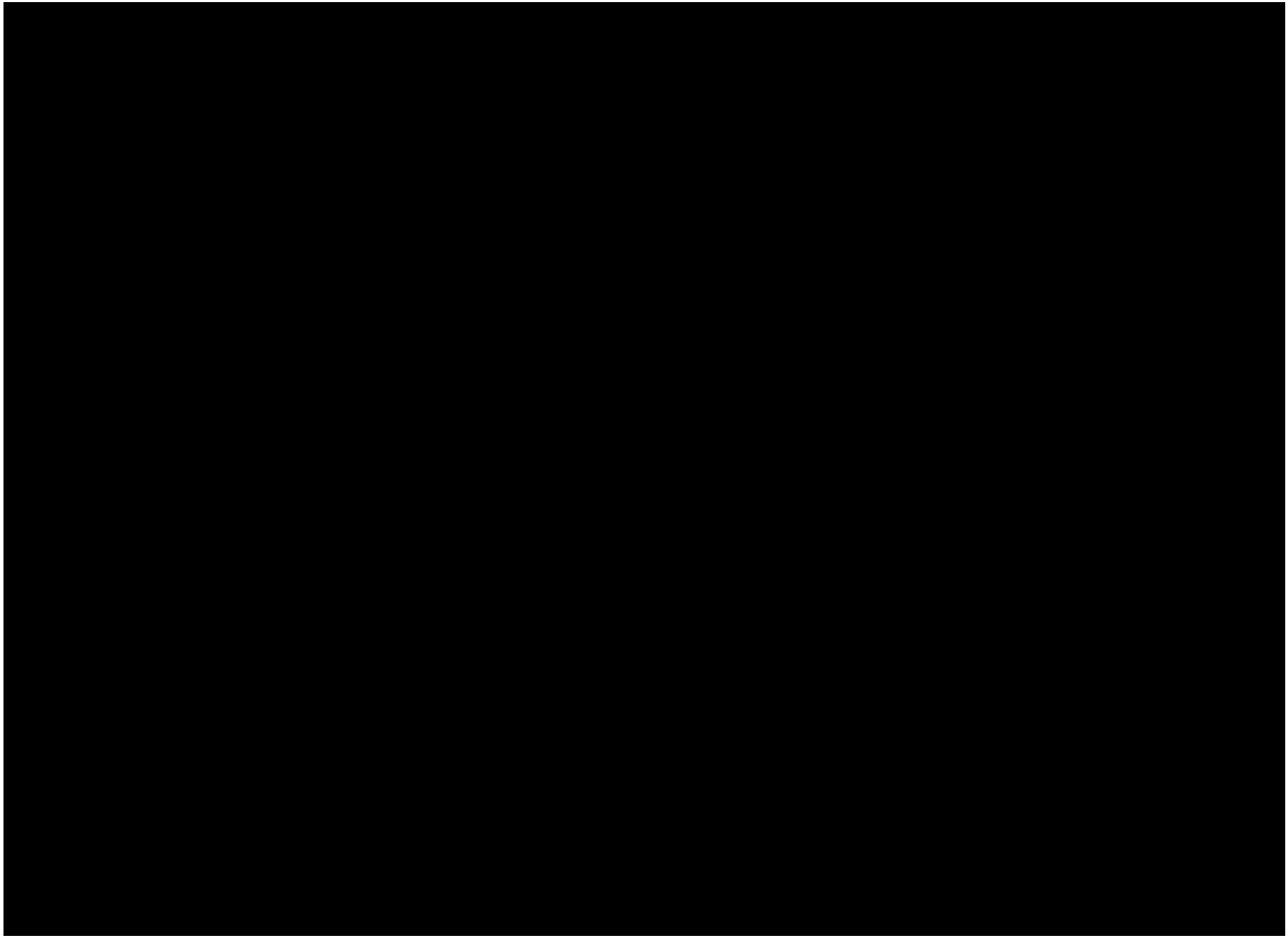


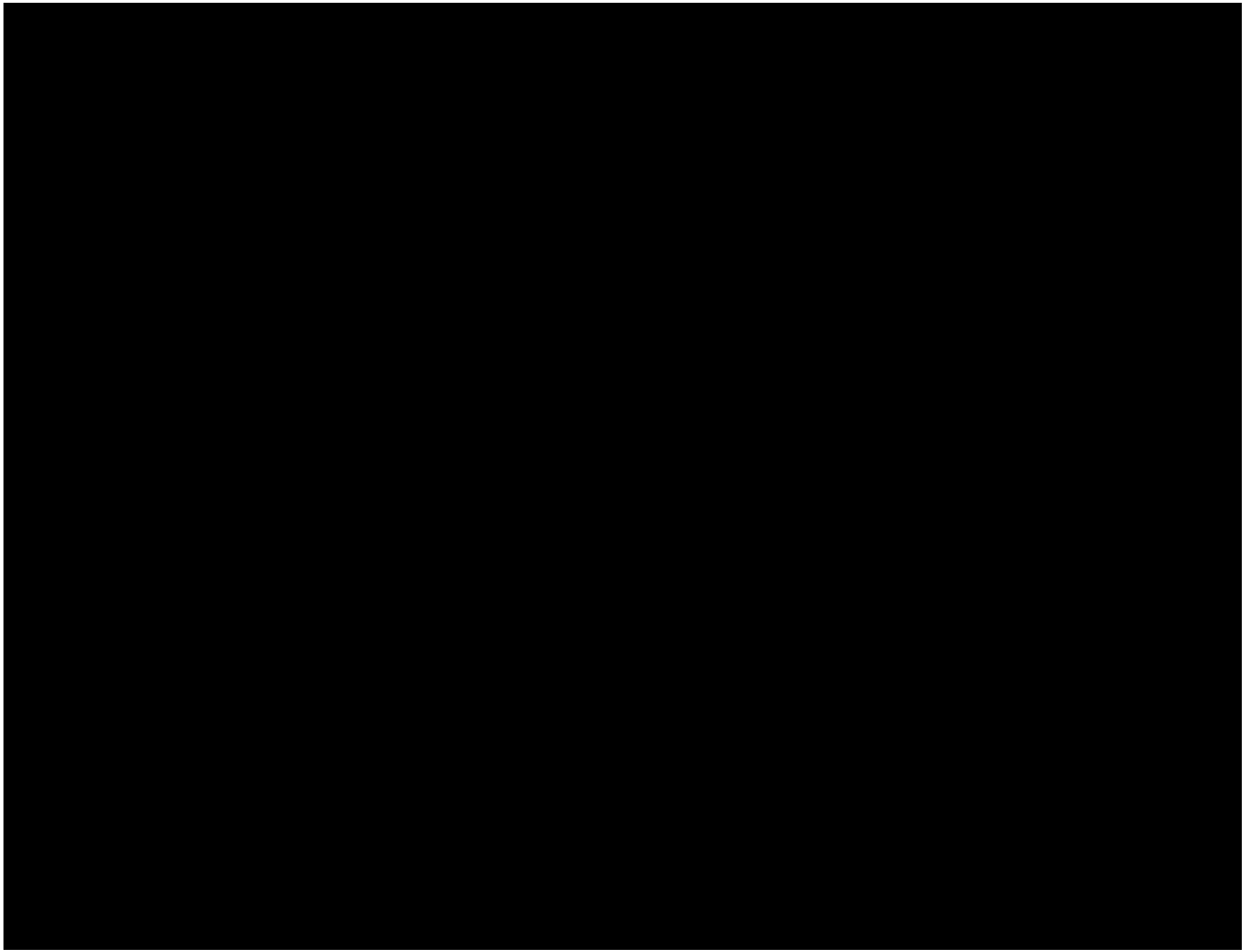


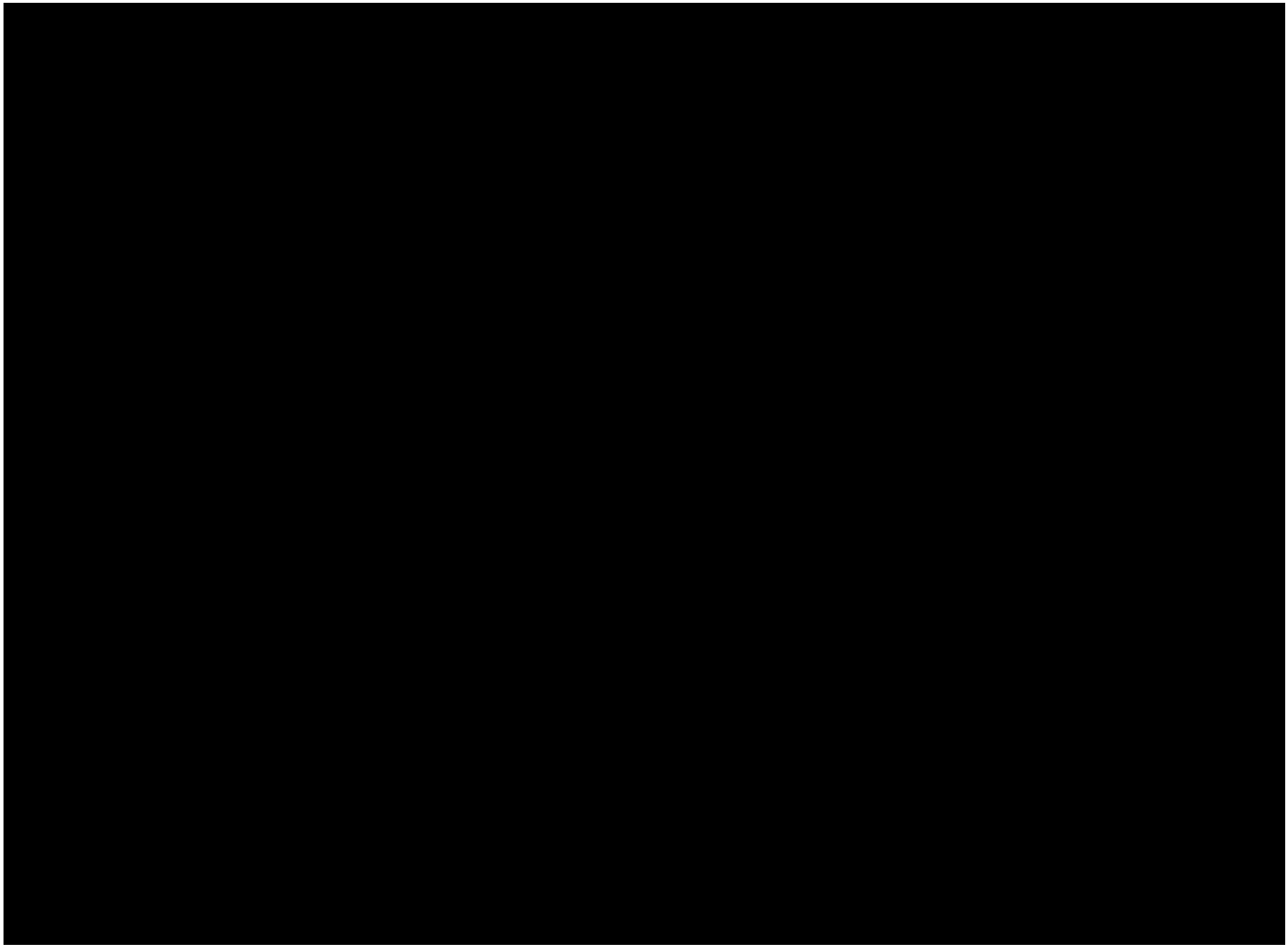


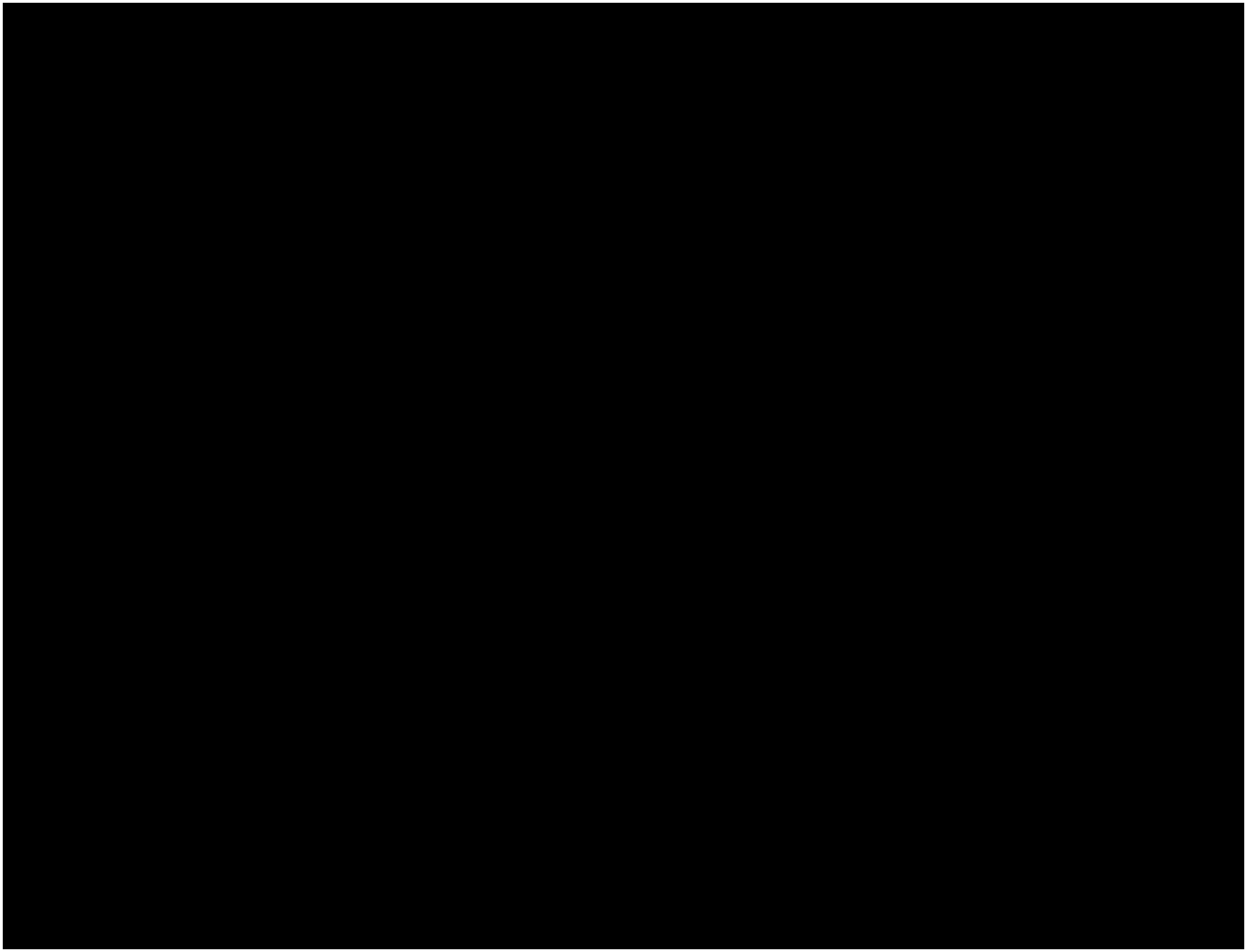


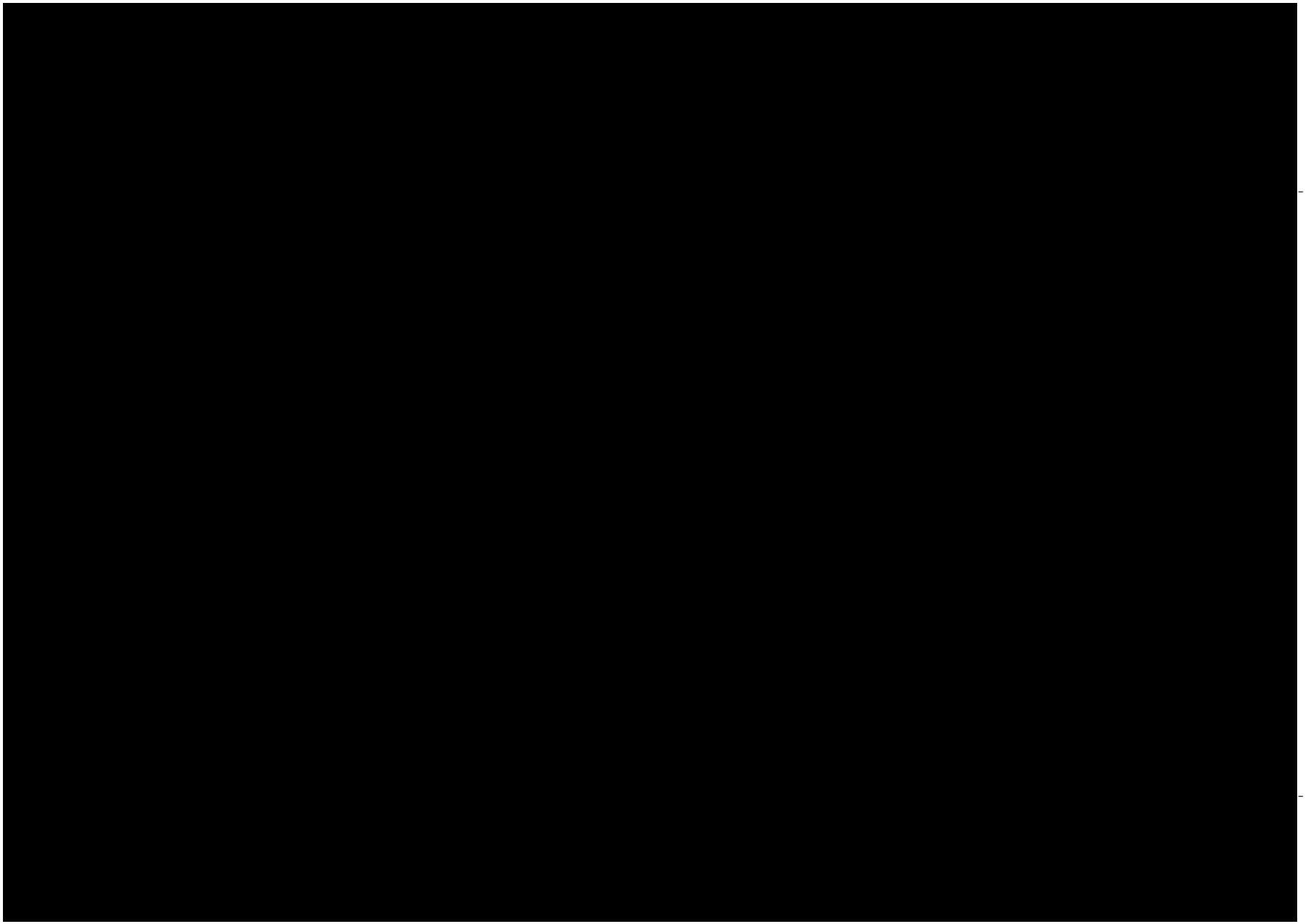












FILE: cp_cal6a.lis, RUN: 31MAR2021 14:53
PROGRAM: /sasdata/un0101/projects/ot/ot2401/06_ia202104u/dev/code/cp_cal6.sas