The New England Journal of Statistics in Data Science Volume 2, 112–119 (2024)
DOI: https://doi.org/10.51387/23-NEJSDS46

Indeterminate Data and Handling for Assessing Diagnostic Performance in Imaging Drug Developments

Sue-Jane WANG

Abstract

In diagnostic imaging drug developments, the imaging scan read data in controlled imaging drug clinical trials includes test positive and test negative. Broadly speaking, the standard of reference data are either presence or absence of a disease or clinical condition. Together, these data are used to assess the diagnostic performance of an investigational imaging drug in a controlled imaging drug clinical trial. For those imaging scan read data that cannot be called positive/negative, the "indeterminate" category is commonly used to cover imaging results that may be considered intermediate, indeterminate, or uninterpretable. Similarly, for those standard of reference data that cannot be categorized into presence/absence including uncollected or unavailable reference standard data, the "indeterminate" category may be used. Historically, little attention has been paid to the indeterminate imaging scan read data as they are generally rare or considered irrelevant though they are related to scanned subjects and can be informative. Subjects lack the standard of reference are simply excluded as such the study only reports the analysis results in subjects with available standard of reference data, known as completer analysis, similar to evaluable subjects seen in controlled trials for drug developments.

To improve diagnostic clinical trial planning, this paper introduces five attributes of an estimand in diagnostic imaging drug clinical trials. The paper then defines the indeterminate data mechanisms and gives examples for each indeterminate mechanism that is specific to the clinical context of a diagnostic imaging drug clinical trial. Several imputation approaches to handling indeterminate data are discussed. Depending on the clinical question of primary interests, indeterminate data may be intercurrent events. The paper ends with discussions on imputations of intercurrent events occurring in indeterminate imaging scan read data and those occurring in indeterminate standard of reference data when encountered in diagnostic imaging clinical trials and provides points to consider of estimands for diagnostic imaging drug developments.

KEYWORDS AND PHRASES: Diagnostic imaging drug clinical trial, Estimand, Indeterminate mechanism, Intercurrent event.

1. INTRODUCTION

Following the official release of ICH E9(R1) addendum on "Estimands and sensitivity analysis in clinical trials" in 2020 [1], the drug sponsors and regulatory agency have gradually learned to actively communicate during the trial planning stage [2, 3, 4, 5]. The goal is to identify what should the key clinical question of interest in a therapeutic clinical trial be, which should reflect the study objective. A controlled clinical trial is planned according to its study objective(s). Intercurrent event(s) occurring after trial initiation that affect either the interpretation or the existing measurements associated with the clinical question of interest should be identified a priori and be addressed in order to precisely define the treatment effect to be estimated in a clinical trial [1].

In diagnostic imaging drug developments, the imaging scan reads of eligible subjects following the administration of an investigational imaging drug are compared against the definitive diagnosis known as standard of truth (SoT) or truth standard, more broadly the standard of reference (SoR) or reference standard. The definitive diagnosis for the presence or absence of a disease or clinical condition under evaluation is determined by histopathology including pathology or biopsy (SoT) or, when histopathology is unavailable, determined by a composite of imaging, associated laboratory measures, and clinical follow-up (SoR).

Sensitivity and specificity are generally the pre-specified co-primary efficacy endpoints, which measure the diagnostic performance of an investigational imaging drug [6, 7]. The sensitivity of a test is defined as the ability of the test to identify subjects with the disease or the clinical condition of interest, which is expressed as the proportion of subjects who truly have the disease who are so identified by the test [6, 7, 8]. The specificity of a test is defined as the ability of the test to identify subjects without the disease or without the clinical condition of interest, which is expressed as the proportion of subjects who truly are disease free who are so identified by the test [6, 7, 9].

When the definitive diagnosis is based on the SoR, sensitivity and specificity are referred to as positive percent agreement (PPA) and negative percent agreement (NPA), respectively. However, the conceptual framework of diagnostic performance described for sensitivity and for specificity is equally applicable for PPA and NPA. Thus, the broadly

used term of SoR as the measure of definitive diagnosis and commonly understood terms of sensitivity and specificity as measures of diagnostic performance will be used throughout the rest of the paper.

In Section 2, the intent-to-image (ITI) set and the modified ITI (mITI) set of the imaging analysis population in a diagnostic imaging drug clinical trial are introduced. Intercurrent events along with necessary attributes in diagnostic imaging drug clinical trials are described in Section 3. For imaging scan read data or SoR data that cannot be accurately determined, known as indeterminate data in diagnostic imaging literature [10, 11, 12], there are at least three types of indeterminate mechanisms. Section 4 gives descriptions of each indeterminate mechanism and provides examples by type of indeterminate mechanism. Indeterminate data may be intercurrent events, which depend on the key clinical question of interest in a diagnostic imaging drug clinical trial. More expanded than therapeutic trials, multiple approaches to handling indeterminate data encountered in diagnostic imaging drug clinical trials are elaborated in Section 5. Discussions on intercurrent events along with points to consider in diagnostic imaging drug clinical trials follow in Section 6.

2. INTENT-TO-IMAGE SET VERSUS MODIFIED INTENT-TO-IMAGE SET

Subjects enrolled in a diagnostic imaging drug clinical trial satisfying the inclusion/exclusion criteria are the intent-to-image (ITI) subjects. In principle, sensitivity and specificity should be estimated from the ITI subjects. However, definitive diagnosis of ITI subjects may not always be collected and some subjects may not be imaged properly or may discontinue from an imaging drug administration due to acute side effect(s) not immediately resolvable. When definitive diagnoses of ITI subjects are not all collected, the estimated sensitivity and specificity or the interpretation of the diagnostic performance relying upon only those subjects whose definitive diagnosis are available will be affected and may be biased. The concern of bias could be mounting if decisions to not collect the definitive diagnosis SoR data are made simply based on the results of the imaging scan read data or the subjective judgments made by investigators or treating physicians who are unblinded to the imaging scan read results.

Subjects receiving the investigational imaging drug are expected to undergo imaging scans at some pre-specified time window based on the half-life of the investigational imaging drug. Instead of the ITI set, i.e., all enrolled eligible subjects, some argue that modified intent-to-image (mITI) set in whom the investigational imaging drug are administered should be the primary diagnostic efficacy analysis set. Consequently, the mITI set may be loosely referred to as the ITI set or all subjects scanned. In therapeutic trials, the modified intent-to-treat (mITT) set are those subjects randomized and received the investigational drug as opposed to the intent-to-treat (ITT) set in those subjects randomized. The ITT principle is carefully articulated in ICH E9 [13], which preserves the randomization for objective assessment of treatment effect in a randomized controlled therapeutic trial. In parallel with therapeutic trials, the mITI set or all subjects scanned is not the ITI set.

3. INTERCURRENT EVENT

Intercurrent events are events occurring after treatment initiation that affect either the interpretation or the existence of the measurements associated with the clinical question of interest (ICH E9(R1)). It is necessary to address intercurrent events when defining the clinical question of interest to precisely define the treatment effect that is to be estimated. The clinical question of primary interest in a diagnostic imaging drug clinical trial can help frame what the estimand should be. In structuring the description of an estimand, the following attributes are proposed in the ICH E9(R1) addendum [1]. For each attribute, the context in a diagnostic imaging drug clinical trial is formulated.

- Population: the population of interest as reflected in the inclusion/exclusion criteria of a given trial. Like the ITT principle or the mITT principle in a therapeutic trial, a diagnostic imaging drug clinical trial considers the intent-to-image ITI or the mITI principle in defining the imaging population.
- Treatments: the specific treatments to be compared. When a diagnostic imaging drug clinical trial uses a parallel-arm study design such as a placebo-controlled trial [14], the principle of specific treatments to be compared following randomization is similar to that in a therapeutic clinical trial. For an intra-subject single-arm controlled design, the comparison of interest is inherently within subject. For instance, it may be of interest to compare images without an imaging drug (pre-image or no contrast) to images after an imaging drug (or with contrast) is administered (e.g., [15]). Other comparisons may be of interest such as comparing to a pre-specified threshold that should be clinically meaningful [16, 17], or to an active comparator in a within-patient crossover design [18].
- Variable: the endpoint of scientific interest. The diagnostic endpoint of scientific interest depends on the clinical context in a specific disease setting under investigation. In defining what constitutes diagnostic efficacy for a subject, the diagnostic endpoint may be measured at subject level or a more granular level, e.g., disease detection at the lesion level, nodal identification at the node level.
- Intercurrent events: an intercurrent event in a diagnostic imaging drug clinical trial may be the lack of SoR data if the scientific question of interest is what the diagnostic performance of an investigational imaging drug is. Strategies on how to account for intercurrent

events should reflect the scientific question of primary interest. A distinguishing feature of a diagnostic imaging drug is that although it is an investigational drug, the drug is to diagnose a subject, but not to treat a subject. The SoR reflects a subject's truth disease status or clinical condition, which should not be affected by the imaging drug test results.

- Summary measure: a summary measure is a variable which provides a basis for a comparison between different treatment conditions. In diagnostic imaging drug clinical trials, the diagnostic efficacy is a summary of accumulated success either at the subject-level or a more granular level that is clinically meaningful.

4. IMAGING SCAN READ DATA AND STANDARD OF REFERENCE DATA

There are two types of diagnostic efficacy data needed to evaluate the co-primary efficacy endpoints of sensitivity and specificity in a diagnostic imaging drug clinical trial. One type is imaging scan read data, i.e., scan read results of the imaging test collected following an investigational imaging drug administration. The other type is standard of reference data, viz., the definitive diagnosis to be determined based on data obtained from histopathology, pathology or biopsy procedure that are generally invasive, or based on other imaging data, lab results or clinical follow-up if histopathology data are unavailable.

To assess the diagnostic efficacy of an investigational imaging drug, it is necessary to make plausible assumptions on the indeterminate data in the imaging scan read data and/or SoR data. Following the three types of missing mechanisms initially proposed by Little and Rubin [19] and adopted in therapeutic trials, below, these missing mechanisms are illustrated in the context of a diagnostic imaging drug clinical trial.

4.1 Indeterminate Mechanism of Imaging Scan Read Data

The imaging scan read results of the diagnostic imaging test are positive, negative, or indeterminate. The term indeterminate could mean intermediate if the imaging results are derived from multiple categories or the threshold used for binarization (positive, negative) is derived from continuous intensity measures, not sure or uninterpretable [11]. An indeterminate imaging scan read may be due to device required for performing the imaging drug test is malfunctioned, imaging quality is sub-optimal, or is related to diagnostic imaging drug safety. Table 1 lists the indeterminate data mechanisms of imaging scan reads, gives descriptions, and provides examples of each indeterminate imaging scan read data mechanism.

Most investigational diagnostic imaging drugs are safe, as the dose exposure level identified in early phase proof-of-concept trials and selected for confirmatory imaging drug trials should be the optimal dose defined as the lowest exposure dose that yields comparable diagnostic efficacy than any higher dose level without the need to risk exposing to a higher dose level. If diagnostic imaging drug scan data cannot be determined, the more frequently seen indeterminate data mechanism is indeterminate at random (IAR), or equivalently, missing at random (MAR) in therapeutic trials. For instance, a subject may not follow the instruction of the imaging procedure such as a subject could not lie still or has moved during an imaging session, which might be age and/or health condition related. The less frequently seen indeterminate data mechanism of imaging scan read data are indeterminate completely at random (ICAR) due to device malfunction or indeterminate not at random (INAR) due to imaging drug induced safety events. The therapeutic counterpart of ICAR is MCAR (missing completely at random) and is MNAR (missing not at random) for INAR.

4.2 Indeterminate Mechanism of SoR Data

The definitive diagnosis to establish the standard of reference includes being present, absent, and indeterminate of a subject's disease status or clinical condition, e.g., Parkinsonian syndromes, cancer metastasis. The definitive diagnosis is known to be invasive. An indeterminate SoR may be due to insufficient amounts of tissues or specimens for histopathology evaluation, difficulty in reaching the surgical cavity for specimen collection, investigator's judgment of subject's health status or medical condition and who is unblinded to subject's imaging scan read data or clinical information, physician's decision of a subject's treatment option or treatment management and who is unblinded to subject's image read data and clinical information. Table 2 lists the indeterminate SoR data mechanisms, gives descriptions, and provides examples of each indeterminate SoR data mechanism.

When the definitive diagnosis (SoR) is indeterminate, it is very unlikely to be indeterminate completely at random (ICAR). The indeterminate data mechanism is mostly associated with either subjects' own tissues or specimens, or investigators or treating physicians who take care of the subjects. Therefore, it is listed as "Unlikely scenario" under examples for ICAR indeterminate data mechanism in Table 2. In a diagnostic imaging drug clinical trial, if investigators or treating physicians are blinded to diagnostic imaging scan read data, the likelihood of the indeterminate SoRs seems tend toward indeterminate at random (IAR). Otherwise, one would have to suspect the likelihood of these indeterminate SoRs would, more likely than not, be indeterminate not at random (INAR) for their being confounded with investigators or treating physicians unblinded to imaging scan read results.

5. INDETERMINATE DATA HANDLING

For data analysis purposes, the efficacy endpoints in diagnostic imaging drug clinical trials are binary measures and

Table 1. Indeterminate mechanisms of imaging scan read data in diagnostic imaging drug clinical trials.

Indeterminate Imaging Scan Read Data Mechanism	Description	Examples
Indeterminate completely at random (ICAR)	The likelihood of the imaging scan read data being indeterminate is unrelated to observed and unobserved variables, i.e., the indeterminate status is unrelated to the specific subject or imaging drug being studied	The device required to perform the imaging scan read following administration of an investigational imaging drug is malfunctioned
Indeterminate at random (IAR)	The likelihood of the imaging scan read data being indeterminate is related to observed variables but not to unobserved variables, i.e., the indeterminate is related to the subject and can be predicted from other data known about that subject	The imaging scan read data are indeterminate due to subjects not following the instruction of the imaging procedure, e.g., could not lie still or have moved during an imaging session, which could be, e.g., age and/or health condition related
Indeterminate not at random (INAR)	The likelihood of the imaging scan read data being indeterminate depends on the unobserved data and is related to the investigational drug and outcome measure	Subject cannot complete the imaging procedure due to imaging administration associated side effect (drug safety related) and early discontinued from the study

Table 2. Indeterminate mechanisms of SoR data in diagnostic imaging drug clinical trials.*

Indeterminate SoR Data Mechanism	Description	Examples
Indeterminate completely at random (ICAR)	The likelihood of the SoR data being indeterminate is unrelated to observed and unobserved variables, i.e., the indeterminate status is unrelated to the specific subjects being studied	Unlikely scenario
Indeterminate at random (IAR)	The likelihood of the SoR data being indeterminate is related to observed variables but not to unobserved variables, i.e., the indeterminate status is related to the subject but can be predicted from other data known about that subject	– SoR data are indeterminate due to insufficient amount of tissue or specimen taken – SoR procedure cannot be comprehensively performed due to location difficulty in obtaining specimens or tissues
Indeterminate not at random (INAR)	The likelihood of the SoR data being indeterminate depends on informative data	– Investigator's judgment of subject's health status or medical condition and who is unblinded to subject's imaging scan read data or clinical information – Physician's decision of a subject's treatment option or treatment management and who is unblinded to subject's imaging scan read data and clinical information

*SoR: Standard of Reference.

the standards of reference SoR are also binary measures. The clinical question of primary interest may be, e.g., (i) "Is the investigational diagnostic imaging drug effective in diagnosing eligible subjects' true disease status or clinical condition who received an investigational imaging drug for image scan test?", (ii) "Is the investigational diagnostic imaging drug effective in diagnosing eligible subjects' true disease status or clinical condition who received the imaging drug with scan results?". The estimand should reflect the clinical question of primary interest.

When the imaging scan read data or the SoR data are categorized as "indeterminate", imputations of indeterminate data may be necessary in assessing the diagnostic efficacy of all subjects scanned following the administration of an investigational diagnostic imaging drug. Depending on the indeterminate mechanism, the imputation approaches can be with limited sophistication ranging from simple probability assignment to tipping point analysis to using complex strategies including modeling, multiple imputation or mimicking clinical risk probabilities in view of harmonized internationally accepted clinical guidelines.

5.1 Random Probability Imputation

If one can assume that the reason for the indeterminate mechanism is ICAR, that is, the reason is completely unrelated to observed variables and unobserved variables, random probability imputation is an option. For indeterminate imaging scan read data, a 50:50 would assign 0.5 probability of being imaging scan positive and 0.5 probability of being imaging scan negative. For indeterminate SoR data, a 50:50 would assign 0.5 probability of being true positive (TP or SoR is classified as present) and 0.5 probability of being true negative (TN or SoR is classified as absent). It is worth noting, also see Table 2, that the scenario of SoR being indeterminate completely at random, ICAR, appears very unlikely.

5.2 Tipping Point Analysis

The tipping point analysis is a variation to random probability imputation. The entire probability spectrum ranging from 0.0 to 1.0 is evaluated. Extending from 50:50 random probability, the tipping point analysis explores the diagnostic performance in either direction, one towards probability of 0.0, the other towards probability of 1.0. That is, the tipping point analysis extends the diagnostic likelihood toward more likely to be TP (e.g., 60:40 with 0.6 probability of being TP and 0.4 probability of being TN, 70:30, 80:20, 90:10 and in the extreme 100:0 assuming 100% probability with no uncertainty or with probability 1.0 of all being TP among those who lack standard of reference) or toward more likely to be TN (e.g., 40:60 with 0.4 probability of being TP and 0.6 probability of being TN, 30:70, 20:80, 10:90 and in the extreme 0:100 with probability 1.0 of all being TN among those who lack standard of reference). A threshold probability is then identified to match with the pre-specified sensitivity threshold and specificity threshold. This threshold probability identifies the probability cutoff that tips the primary analysis result to be in favor of the investigational diagnostic imaging drug.

5.3 Multiple Imputation

Instead of a single imputed value for filling in an indeterminate data point, the idea of multiple imputation is to create multiple imputed datasets and analyze each of them using some statistical method/model to provide the estimates. Then, the results from all the datasets are pooled to derive the standard errors [19]. Multiple imputation can preserve the relationship between variables in the data, meanwhile, simultaneously consider the uncertainty about the relationships between variables in the data [20].

5.4 Univariate Risk Imputation

For a disease involving prognostic or risk factors, it is possible to impute indeterminate data considering one risk factor at a time. Collectively, the imputed diagnostic performance among one risk factor, two risk factors, etc., can serve as sensitivity analyses to explore the range of possible diagnostic performance and whether there is a risk factor playing a dominate role in the interpretation of the imputed diagnostic performance. The univariate risk imputation can be model-based, risk probability-based with defined risk probability supported by literature or using an ad-hoc rule that is either pre-specified or post-specified as exploratory analyses.

5.5 Multivariate Model-Based Imputation

To diagnose if a subject has the clinical condition or the disease of interest after receiving the investigational imaging drug, pre-specified prognostic or risk factors of the disease can be incorporated in a statistical imputation model with outcome being the probability of disease or clinical condition being present. The model-based imputation yields a probability of disease being present or clinical condition being present, adjusting for subject's prognostic or risk factors. One may be interested in exploring a range of models that are either a subset of or expanded from the pre-specified statistical model that consists of specific prognostic or risk factors; the pre-specified model is the primary imputation model.

5.6 Multiple Risk Probability-Based Imputation

Every subject with indeterminate data will be assigned a probability based on the prognostic or clinical risk factors. Using lack of SoR as an example, the probabilities are pre-specified based on risk probability category(ies) determined by independent experts (for instance, three independent readers) or by consensus reads determined from among experts who are blinded to the imaging scan read results, or by well-established clinical risk factor guideline, for instance, NCCN clinical practice guidelines in oncology treatment by cancer type (https://www.nccn.org/guidelines/category_1) [21].

6. DISCUSSION

A diagnostic imaging drug clinical trial aims to assess if an investigational imaging drug has an acceptable diagnostic efficacy in all subjects who received the imaging drug. Conclusions drawn from diagnostic imaging drug clinical trials with indeterminate data can vary depending on the assumptions of indeterminate mechanism made and the analytical approach chosen. The indeterminate data mechanisms are laid out in Table 1 for imaging scan read data and in Table 2 for SoR data. Examples by indeterminate data mechanism that may occur in diagnostic imaging drug clinical trials are also provided. Noteworthy, some of those indeterminate data that will affect either the interpretation or the existence of the measurements associated with the primary diagnostic efficacy assessment in view of the clinical question of primary interest are *intercurrent events*.

- Handling of intercurrent event occurring in indeterminate imaging scan read data

Historically, little attention has been paid to the indeterminate imaging scan read data that are usually rare or ignored in the analysis, although they are related to scanned subjects and can be informative. There are multiple ways to analyze data with missingness in therapeutic trials, e.g., [22]. There are also multiple approaches to analyze indeterminateness in diagnostic imaging drug clinical trials. Several approaches are illustrated in Section 4.

In rare situation when the acute side effect cannot be resolved in a short time and the number of subjects experiencing the acute side effect cannot be ignored, the acute side effect related to the administration of the investigational imaging drug would be an *intercurrent event*. If a diagnostic imaging drug clinical trial does not early terminate due to acute side effect being not resolvable, the handling of acute side effect as an *intercurrent event* would be important and the primary imputation strategy should be pre-specified in view of the clinical question of primary interest during imaging drug trial planning.

In another situation when the SoR is not indeterminate but the imaging scan read result is indeterminate, random probability imputation can be considered if there is no imaging safety and efficacy concerns. Tipping point analyses generally serve as supportive analyses. The cutoff thresholds from tipping point analyses allow further understanding of the potential benefit-risk tradeoffs between sensitivity and specificity.

- Handling of intercurrent event occurring in indeterminate standard of reference data

In the case that lack of reference standard is an *intercurrent event* with an indeterminate mechanism of SoR data being not completely at random, this event occurrence prevents from proper and accurate assessment of the diagnostic efficacy endpoints (sensitivity and specificity) and thus cannot be ignored. Several imputation approaches to handling indeterminate SoR data are provided. If the reason of indeterminate mechanism is related to observed variable but not to unobserved variables (IAR) or is dependent on the unobserved data (INAR), multivariate model-based imputation, multiple risk probability-based imputation, and multiple imputation version of the above two approaches are plausible imputation approaches to lack of SoR intercurrent event. Since reasons for lack of SoR mainly tie to subject's tissue sample availability, it is unlikely to come up with a scenario where lack of standard reference data can be considered as indeterminate completely at random.

Often, however, radiologists and clinicians are more used to risk probability-based imputation. Risk probability that is pre-specified seems to be easier to grasp than risk probability that is derived from a statistical model, although each approach relies on its specific assumptions. Moreover, multiple risk probability-based imputation allows adjustment of the estimated prevalence through imputation. This approach is particularly informative when disease prevalence of the intent-to-image or modified intent-to-image subjects can be obtained through published literature. The multiple risk probability-based imputation is also suitable for exploring the diagnostic performance in a practical range of disease prevalence reported in medical literature.

- Points to consider of estimands in diagnostic imaging drug development

The imaging drug scan test results need to be compared against the SoR to conclude if the imaging drug scan test after administration of an investigational imaging drug can be used to replace the SoR for clinical use upon approval. Based on the study protocol, SoR should be collected. Clearly, when there is a lack of SoR and the reason being investigators or treating physicians are unblinded to and are influenced by the imaging drug scan read results obtained following the administration of the investigational imaging drug, the estimates of sensitivity and specificity would be biased and the interpretation of the estimands would be affected. Similarly, diagnostic imaging drug scan data should be collected. In cases where indeterminate data mechanism of an investigational imaging drug scan may affect the estimands and estimates of diagnostic performance, handling of such indeterminate imaging drug scan data would be important.

ICH E9(R1) addendum [1] advocates a framework to align planning, design, conduct, analysis, and interpretation. Clear trial objective should be translated into key clinical question of interest to facilitate defining a suitable estimand. In this spirit, a key clinical question of interest in a diagnostic imaging drug clinical trial can be "Is the investigational imaging drug effective, measured by sensitivity and specificity, in diagnosing the disease status of all subjects who received the imaging drug?"

Furthermore, one should carefully define the diagnostic efficacy of interest in a way that determines both *the population of subjects to be included in the estimation of diagnostic efficacy* and *the observations from each subject to be included in the analysis* considering the occurrence of intercurrent events. For the former, population of subjects may be those eligible subjects who received the imaging drug. For the latter, observations of each subject would be imaging scan read data and their SoR data including indeterminate imaging scan read data and indeterminate SoR data, in addition to subject's baseline characteristics. An estimand of a diagnostic imaging drug clinical trial should define the target of estimation for a particular trial objective, viz., the estimators of sensitivity and specificity in all eligible subjects who received the imaging drug. Then, intercurrent events considering the primary clinical question can be identified a priori, see, e.g., Tables 1 and 2.

Pre-specification of anticipated intercurrent event(s) in view of the key clinical question of interest and a suitable

primary method of estimation including the primary imputation approach or strategy to indeterminate data handling based on their indeterminate mechanisms are critical to the success of a diagnostic imaging drug clinical trial. In parallel with therapeutics trials, it is of interest to articulate strategies, such as "imaging policy", "while-on-imaging policy", "hypothetical strategy", etc. in view of selecting approaches for handling intercurrent events in diagnostic imaging drug clinical trials in a future article. To our knowledge, this paper is the first that attempts the topic of estimands and intercurrent events in diagnostic imaging drug clinical trials following the release of ICH E9(R1) [1].

It is expected that the iterative dialogues between the imaging drug sponsor and the regulator will enhance the interactions when discussing the suitability of imaging drug trial designs, and the interpretation of imaging trial results. In this paper, we articulate intercurrent events one may be faced with when evaluating sensitivity and specificity in diagnostic imaging drug clinical trials for diagnostic imaging drug developments. Other imaging drug trials could be, e.g., for intra-operative imaging drug developments [23], theranostics that co-develop an imaging drug and a therapeutic drug [24, 25], or, of a different primary efficacy endpoint with different clinical question of interest [17, 26]. The framework articulated in ICH E9(R1) and in this article can assist imaging drug sponsors in planning imaging drug clinical trials and assist regulators in their reviews of the imaging drug clinical trial statistical analysis plans during the planning of investigational imaging drug developments and the interpretation of trial results upon submissions for new drug applications.

ACKNOWLEDGEMENTS

The author would like to thank the regulatory imaging statistics team and imaging clinical teams for their inputs on indeterminate data reasoning of diagnostic imaging drug clinical trials in multiple clinical settings.

DISCLAIMER

This article reflects the views of the author and should not be construed to represent the views or policies of the U.S. Food and Drug Administration.

Accepted 13 April 2023

REFERENCES

[1] ICH E9 (R1) addendum. Estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials (2020). https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e9-r1-addendum-estimands-sensitivity-analysis-clinical-trials-guideline-statistical-principles_en.pdf.

[2] Aroda, V. R., Saugstrup, T., Buse, J. B., Donsmark, M., Zacho, J. and Davies, M. J. (2019). Incorporating and interpreting regulatory guidance on estimands in diabetes clinical trials: the PIONEER 1 randomized clinical trial as an example. *Diabetes, Obesity and Metabolism* **21**(10) 2203–2210.

[3] Darken, P., Nyberg, J., Ballal, S. and Wright, D. (2020). The attributable estimand: a new approach to account for intercurrent events. *Pharmaceutical Statistics* **19**(5) 626–635.

[4] Wharton, S., Astrup, A., Endahl, L., Lean, M. E. J., Satylganova, A., Skogaard, D., Wadden, T. A. and Wilding, J. P. H. (2021). Estimating and reporting treatment effects in clinical trials for weight management: using estimands to interpret effects of intercurrent events and missing data. *International Journal of Obesity* **45**(5) 923–933.

[5] Yassi, N., Hayward, K. S., Campbell, B. C. V. and Churilov, L. (2021). Use of the Estimand framework to manage the disruptive effects of COVID-19 on stroke clinical trials. *Stroke* **52**(11) 3739–3747.

[6] Zhou, X. H., Obuchowski, N. A. and McClish, D. K. (2002). *Statistical methods in diagnostic medicine*. John Wiley and Sons, New York. https://doi.org/10.1002/9780470317082. MR1915698

[7] Pepe, M. S. (2003). *The statistical evaluation of medical tests for classification and prediction*. Oxford University Press, New York. MR2260483

[8] Guyatt, G., Drummond, R., Meade, M. O. and Cook, D. J. (eds). (2008). *Users' guides to the medical literature: a manual for evidence-based clinical practice*. 2nd ed. McGraw Hill, New York.

[9] Haynes, R. B., Sackett, D. L., Guyatt, G. H. and Tugwell, P. (2005). *Clinical epidemiology: how to do clinical practice research*. 3rd ed. Lippincott Williams and Wilkins, Philadelphia, PA.

[10] Simel, D. L., Feussner, J. R., Delong, E. R. and Matchar, D. B. (1987). Intermediate, indeterminate, and uninterpretable diagnostic test results. *Medical Decision Making* **7**(2) 107–114.

[11] Massion, P. P. and Walker, R. C. (2014). Indeterminate pulmonary nodules: risk for having or for developing lung cancer? *Cancer prevention research* **7**(12) 1173–1178.

[12] Vetter, T. R., Schober, P. and Mascha, E. J. (2018). Diagnostic testing and decision-making: beauty is not just in the eye of the beholder. *Anesthesia & Analgesia* **127**(4) 1085–1091.

[13] Ich E9 statistical principles for clinical trials (1998). https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-9-statistical-principles-clinical-trials-step-5_en.pdf.

[14] Repici, A., Wallace, M. B., East, J. E., Sharma, P., Ramirez, F. C., Bruining, D. H., Young, M., Gatof, D., Canto, M. I. M., Marcon, N., Cannizzoro, R., Kiesslich, R., Rutter, M., Dekker, E., Siersema, P. D., Spaander, M., Kupcinskas, L., Jonaitis, L., Bisschops, R., Radaelli, F., Bhandari, P., Wilson, A., Early, D., Gupta, N., Vieth, M., Lauwers, G. Y., Rossini, M. and Hassan, C. (2019). Efficacy of per-oral methylene blue formulation for screening colonoscopy. *Gastroenterology* **156**(8) 2198–2207.e1.

[15] Gadavist (gadobutrol) injection, for intravenous use. Highlights of prescribing information. https://www.accessdata.fda.gov/drugsatfda_docs/label/2020/201277s018lbl.pdf. Accessed on November 17, 2021.

[16] The NDA multi-disciplinary review and evaluation: NDA 212643, Gallium Ga 68 PSMA-11 injection. https://www.accessdata.fda.gov/drugsatfda_docs/nda/2020/212643Orig1s000MultidisciplineR.pdf. Accessed on November 17, 2021.

[17] Wang, S. J. 2023. Diagnostic test evaluation book chapter in "Translational Radiation Oncology: Handbook for Designing and Conducting Clinical and Translational Research" Editors: Adam EM Eltorai, Jeff Bakal, David E Wazer, Daniel Wonsoon Kim Elsevier Publisher. (Under gallery proof to be published in 2023). MR2828487

[18] Vaneckova, M., Herman, M., Smith, M. P., Mechl, M., Maravilla, K. R., Weichet, J., Spampinato, M. V., Žižka, J., Wippold, F. J. II, Baima, J. J., Babbel, R., Bultmann, E., Huang, R. Y., Buhk, J. H., Bonafe, A., Colosimo, B. C., Liu, S., Kirchin, M. A., Shen, N., Pirovano, G. and Spinazzi, A. (2015). The benefits of high relaxivity for brain tumor imaging: results of a multicenter intraindividual crossover comparison of gadobenate

dimeglumine with gadoterate meglumine (The BENEFIT study). *American Journal of Neuroradiology* **36**(9) 1589–1598.

[19] Little, R. J. A. and Rubin, D. B. (2019). *Statistical analysis with missing data*. 3rd ed. *Wiley Series in Probability and Statistics*. John Wiley and Sons, New York, NY. https://doi.org/10.1002/9781119013563. MR1925014

[20] Van Buuren, S. and Groothuis-Oudshoorn, K. (2011). Multivariate imputation by chained equations in R. *Journal of Statistical Software* **45**(3) 1–67. https://www.jstatsoft.org/v45/i03/.

[21] National comprehensive cancer network (NCCN) clinical practice guidelines in oncology. https://www.nccn.org/guidelines/category_1. Accessed on January 15, 2022.

[22] Little, R. J., D'Agostino, R., Cohen, M. L., Dickersin, K., Emerson, S. S., Farrar, J. T., Frangakis, C., Hogan, J. W., Molenberghs, G., Murphy, S. A., Neaton, J. D., Rotnitzky, A., Scharfstein, D., Shih, W. J., Siegel, J. P. and Stern, H. (2012). The prevention and treatment of missing data in clinical trials. *New England Journal of Medicine* **367**(14) 1355–1360.

[23] Lymphoseek (technetium Tc 99m tilmanocept) injection, for subcutaneous, intradermal, subareolar, or peritumoral use. Highlights of prescribing information. https://www.accessdata.fda.gov/drugsatfda_docs/label/2021/202207s012lbl.pdf. Accessed on November 17, 2021.

[24] Wang, S. J. (2020). Recent advances in clinical trial design considerations in Thera "nostics". *Contemporary Clinical Trials* **96** 106100. https://doi.org/10.1016/j.cct.2020.106100.

[25] Wang, S. J., Fotenos, A., Masters, S. C. and Marzella, L. (2022). Theranostics approach in drug development: is there study efficiency when the prevalence of the molecular target is very high? *Theranostics* **12**(7) 3079–3083. https://www.thno.org/v12p3079.htm. https://doi.org/10.7150/thno.71788.

[26] Pylarify (piflufolastat F18) injection, for intravenous use. Highlights of prescribing information. https://www.accessdata.fda.gov/drugsatfda_docs/label/2021/214793s000lbl.pdf. Accessed on November 17, 2021.

Sue-Jane Wang. Division of Biometrics I, Office of Biostatistics, Office of Translational Sciences, Center for Drug Evaluation and Research, U.S. Food and Drug Administration, 10903 New Hampshire Ave. Building 21, Room 3506, Silver Spring, MD 20993-0002, USA.
E-mail address: suejane.wang@fda.hhs.gov