Published in final edited form as:

J Empir Res Hum Res Ethics. 2019 April ; 14(2): 95–106. doi:10.1177/1556264619831888.

Measuring IRB Regulatory Compliance: Development, Testing and Use of the National Cancer Institute StART Tool

Lisa Rooney¹, Laura Covington², Andrea Dedier², and Birdena Samuel²

¹Frederick National Laboratory for Cancer Research sponsored by the National Cancer Institute (USA)

²The Emmes Corporation, Rockville, MD

Keywords

IRB; stipulations; regulatory; compliance; tool; clinical trials

Introduction

Both United States based Institutional Review Boards (IRBs) and international Research Ethics Committees (RECs), review, approve and oversee Department of Health and Human Services (HHS) funded as well as Food and Drug Administration (FDA) regulated research. These entities are responsible for protecting the rights and welfare of people who participate in HHS funded/FDA regulated biomedical and social/behavioral research by ensuring that non-exempt human subjects research protocols satisfy federal regulations governing study approval criteria (21 CFR 56.111 and 45 CFR 46.111) throughout the entire course of such study protocols. IRBs have been criticized for overstepping their authority by requiring study protocols to meet requirements or satisfy stipulations that go beyond regulatory requirements (Gunsalus et al., 2006; White, 2007). The practical implications are that IRB overreach (Gunsalus et al., 2006; White, 2007) or wordsmithing when stipulating changes to research protocols may do little to enhance the protection of study participants and can cause delays in conducting important research (Whitney, 2016). Moreover, because federal regulatory criteria lends itself to multiple and varying interpretations, U.S. IRBs differ in their application of the federal regulations (Abbott & Grady, 2011; Nicholls et al., 2015) and may fail to address all required Common Rule criteria (Lidz et al., 2012). Further, avoiding the practices of overreach and wordsmithing may alleviate common criticisms and improve efficiency.

The National Cancer Institute's (NCI) Cancer Prevention and Control (CPC) Central Institutional Review Board (CIRB) is the fourth and most recently constituted NCI CIRB. It was established in December 2014 and began reviewing studies in February 2015 (Massett et al., 2018). The CPC CIRB is responsible for review, approval and oversight of a broad portfolio of research funded by NCI's Division of Cancer Prevention (DCP). After 1 year of

Corresponding author: Lisa Rooney, 9609 Medical Center Drive, Rockville, MD 20850; phone: 301-933-3913, lisa.rooney@nih.gov.

Disclaimers [if any]: none

experience, an analysis found that CPC CIRB took longer to approve studies and issued more stipulations compared to NCI's existing and more mature boards (Masset et al., 2018). An assessment was needed to determine the extent to which the new board was adhering to regulatory approval criteria when rendering stipulations, but a literature search revealed no appropriate IRB/REC assessment tools existed. A new tool to assess the CPC CIRB's regulatory compliance was thus developed, tested and effectively used to evaluate and categorize the content of stipulations rendered over a 21-month period.

Methods

Literature Search

A literature search was conducted to identify tools designed to evaluate IRB/REC adherence to the regulatory decision criteria. The search revealed that Lidz (Lidz et al., 2012) used audio recordings to determine whether the IRB deliberations of 10 academic medical centers appropriately considered all applicable approval criteria outlined in 45 CFR 46. Other efforts have evaluated IRB decision letters to: (a) identify the areas of research most likely to receive stipulations (Sansone, McDonald, Hanley, Sellbom, & Gaither, 2004); (b) identify stipulations related to scientific issues (Angell, Bryman, Ashcroft, & Dixon-Woods, 2008); (c) identify stipulations made by IRBs at a single site (Adams et al., 2014; Blackwood et al., 2015; Olsen & Mahrenholz, 2000; Sansone et al., 2004); (d) categorize the types of changes requested (Kent, 1999); and (e) assess whether justifications for stipulations were included in the letter (Clapp, Gleason, & Joffe, 2017). But there were no published articles addressing the use of tools specifically designed to assess adherence to 45 CFR 46.111, the Common Rule regulatory approval criteria, or 21 CFR 56.111, FDA regulatory approval criteria.

Development of the NCI StART Tool

After the literature search finding, an NCI decision was made to create a pilot tool for IRB assessment purposes. NCI asked a regulatory consultant and an NCI employee to create such a pilot tool. The Tool ultimately created for this purpose was the NCI CIRB Stipulation Analysis Review Tool (NCI StART). NCI StART was developed after taking into account the IRB regulatory approval criteria outlined in 21 CFR 56.111 and 45 CFR 46.111, review of relevant literature, and the professional experience of the two individuals charged with the development of the Tool. Thus, the NCI StART categories to which stipulations were ultimately assigned were derived from the above-referenced resources.

NCI StART was designed to assign stipulations into one of two distinct categories as follows. (See the Supplemental Digital Content for the complete NCI StART tool document.)

Stipulations Related to Approval Criteria

The following NCI StART categories were generated to align with the IRB approval criteria found at 21 CFR 56.111 and 45 CFR 46.111:

- Risks to Subjects Minimized;

- Risks Reasonable in Relation to Anticipated Benefits; Equitable Selection of Subjects;
- Informed Consent;
- Waiver or Alteration of Informed Consent Process or Documentation of Informed Consent;
- Data Monitoring to Ensure Subject Safety; Privacy and Confidentiality; and
- Vulnerable Subjects.

Sub-categories were added to some of the categories noted above in order to reduce subjective uncertainty and to provide increased reliability in correctly assigning the wide range of content found in the stipulations. For example, the category “Equitable Selection of Subjects” was assigned to the sub-categories of Inclusion/Exclusion Criteria and Screening/Recruitment/Enrollment. Occasionally, IRB letters were found to contain an administrative artifact of “compound stipulations,” meaning that two discrete items were combined into a single stipulation rather than presented as two separate stipulations. The coding team determined that these would be coded to the individual stipulation level rather than as a general non-descript compound stipulation to provide the user of the Tool with insight regarding each of the individual stipulations captured in the compound stipulation.

Non-IRB Approval Criteria Categories

The Non-IRB Approval Criteria represent categories of stipulations that do not specifically align with the regulatory decision criteria noted above. The three categories are editorial, inconsistencies, and non-categorizable.

The editorial category captures those stipulations that require changes to text to improve clarity or content. This category includes stipulations that correct spelling, grammar, typing, and formatting errors to improve clarity or content. These stipulations may be directed toward the protocol or the informed consent document.

The inconsistencies category of stipulations may be associated with either the application, protocol, or consent documents and indicate inconsistencies either between documents or within documents.

The non-categorizable stipulations cannot be assigned to any of the previous categories and may represent items commonly associated with IRB overreach (mandating stipulations that have no reasonable relationship to regulatory requirements) and wordsmithing (requiring changes to text based on preference without improving clarity or content). Thus, only those stipulations that could not be coded to any other category or subcategory, or that required changes based on preference without improving clarity or content, were coded to this category.

See Table 1 for a list of category explanations, sub-category distinctions (where applicable) and corresponding example stipulations.

Testing of the NCI StART Tool

Following the initial design of NCI StART, the Tool was used by a select number of users of NCI StART, i.e., coders, to ascertain whether such coders, when assessing the same CPC CIRB determination letters, would categorize CPC stipulations contained in CPC CIRB determination letters similarly or whether such testing would reveal that certain NCI StART categories would need further revision or require additional operational definitions and/or specific examples to reach appropriate agreement among NCI StART users. The testing process is explained below.

Data Source

Certain CIRB decision letters generated in 2015 with determinations of “approved pending modification” (APM) or “tabled” were used for the NCI StART development process, as described below. Each determination letter was organized in a similar fashion with individual stipulations numbered sequentially according to their location (Protocol, Consent, Application).

Ten CPC CIRB studies that received either an APM or tabled decision letters during 2015 were selected for coding during Round 1. The same 10 CPC CIRB studies coded during Round 1 were re-coded during Round 2. Three of the 10 protocols previously coded in Round 1 were re-coded during Round 3; three different protocols of the 10 protocols initially coded in Round 1 were re-coded during Round 4; and another three protocols of the 10 protocols initially coded in Round 1 were re-coded during Round 5. Three new CPC CIRB studies were selected for coding during Round 6. A sample of letters drawn from the other three NCI CIRBs—Adult Late Phase Emphasis CIRB (LPE), Pediatric CIRB (Peds), and the Early Phase Emphasis CIRB (EPE)—were coded for Rounds 7 and 8.

As noted above, the coders re-coded a subset of the original 10 selected studies during rounds 2 – 5. Such recoding was undertaken to internally test improvements to the Tool, e.g., revising categories, adding operational definitions, specific examples, to determine whether such improvements would result in the coders remaining consistent in their own assessment of stipulations from previous reviews, with increased confidence and less subjective uncertainty.

Coders

A total of four coders and an arbitrator were involved in the NCI StART testing process. The four coders consisted of a HRPP regulatory consultant and former OHRP compliance oversight coordinator, an NCI employee responsible for oversight of the NCI CPC CIRB and two contractors who are involved in the day-to-day quality review and oversight of the NCI CIRB project. The arbitrator was the quality manager for the NCI CIRB project.

Initially, the review of CPC CIRB stipulations using NCI StART was limited to two coders because the initial coding rounds, i.e., rounds one, two and three, were to pilot test the newly developed NCI StART. The purpose of the pilot testing phase was to identify areas within NCI StART that required refinement before introducing NCI StART to the more expansive coding team. Of note, this pilot testing phase revealed that certain NCI StART categories

needed to be revised while other categories required operational definitions and specific examples to decrease overlap and improve discriminate ability. Once overlap and subjectivity was decreased and concordance between two coders was acceptable, the coding team was expanded to four coders using an updated version of StART. Following the third round, the coding team expanded to include two additional coders and an arbitrator.

Review Process

Eight rounds of review were conducted. For each round, coders were assigned the same board determination letters, and each coded all the stipulations in the letter. The coders independently reviewed the letters and recorded results into a templated worksheet.

After each round of coding, the coders compared their individual coding results and discussed differences in results. Following these discussions, a “friendly reconciliation” took place and individual results were revised if agreement was reached. A reconciled concordance value was then determined. These discussions served as an important part of the process and resulted in a new StART version that was used for each subsequent round of coding. The iterative changes to StART included: creation of operational definitions; creation of sub-categories; revisions to the categories and subcategories; and addition of specific examples.

Data Collection and Analysis

The individual results were transcribed from the reviewer worksheet into an Excel spreadsheet. The Excel data fields included: 1) the numeric code assigned to each StART category and sub-category; 2) protocol number; 3) protocol review date(s); and 4) unique stipulation identifier. Each stipulation was identified by its location (Protocol = P; Consent = C; Application = A) and its numerical identifier in the letter. For example, in Protocol “A”, reviewed on May 1, 2015, “P3” is the code assigned to the third stipulation in the protocol section of the determination letter. Stipulation P3 is then associated with the relevant StART code.

Each coder assigned an NCI StART code to each stipulation. Concordance rates were used to assess the frequency with which the coders assigned the same category to a stipulation. StART categories were then ranked by frequency of discordance and by the combinations of discordant values within each category to identify significant areas of coder variability. Inter-rater reliability for the four coders was calculated using the Krippendorff methodology (Krippendorff, 1970). After the eighth and final round of testing, 72 stipulations were selected randomly for individual test-retest verification to establish intra-rater reliability. The goal of this exercise was to ensure that the coders were consistent in their own assessment of stipulations from previous reviews. For each coder, the re-test value was compared to the coders’ original and reconciled values. Overall, there was a >90% match on the test/re-test round of coding. This shows that the coders were consistent with their previous coding of stipulations.

After achieving satisfactory inter- and intra-rater reliability among and between the coders in October 2016, the final version of the tool was used to code a new set of CPC CIRB determination letters. Partial data was shared with CPC CIRB board members at their

December 2016 in-person Education Day. Board members were appreciative of the feedback and provided suggestions for further refinement of the tool, which was undertaken in early 2017.

Table 2 shows the reconciled concordance rates and inter-rater reliability for the eight rounds of testing. Round 1 of testing included 10 protocols that resulted in 21 determination letters and 299 stipulations. The overall concordance rate (the percentage of stipulations with agreement between the two coders) was 67%. To better understand the variation between the two coders, the StART categories were ranked by frequency of discordance: six categories accounted for 84% of the discordance: clarity; research design; risk exposure; clarify to better understand research, consent, or both; study instruments/procedures; and consent elements. Within each of these six categories, the combinations of discordance were evaluated to identify patterns of selection that accounted for significant coding variance. For example, the coders frequently would have discordance for a stipulation and code it as either clarity or clarify to better understand research. Five such combinations were identified. After discussion and modification, 50 stipulations that represented the most frequently occurring discordant categories were identified for retest. This second look improved the concordance from 67% to 72%. The exercise clarified that some categories needed revision to minimize overlap and improve the coder's ability to pick the right code. For example, the categories "clarity" and "clarify to better understand research, consent, or both" were difficult to assign reliably. As such, operational definitions and specific examples were added to the coding tool. Despite these changes, repeat coding of the original 299 stipulations resulted in a decrease in concordance to 49% in Round 2. However, now only three categories accounted for 78% of the discordance. Following modifications to StART, three of the original 10 studies were re-coded for Round 3 and the overall concordance rate between the two coders increased to 70%. The categories of "Editorial" and "Quality" continued to be challenging to define clearly and code consistently. However, the previously discordant areas of "Research Design" verses "Study Instruments" improved with a decrease from 30% to 7% discordance.

The stipulations in Rounds 4 through 8 were evaluated by four coders. By the end of Round 6, the four coders achieved Krippendorff alpha (Krippendorff, 1970) of 0.739 ("substantial"). However, the alpha dropped significantly in Rounds 7a and 7b when the stipulations issued by two different NCI CIRB boards were evaluated. Variances were identified and discussed by the group, leading to further refinements to problematic areas in the tool. The next round of coding (Round 8) also reviewed stipulations issued by a different board (Adult CIRB - Early Phase Emphasis), and a Krippendorff alpha of 0.869 ("near perfect") was achieved. Though the concordance agreement was not 100%, rates above 80% were established in the final rounds of review. One round of test-retest was conducted with each coder to establish intra-rater reliability. Each of the four reviewers scored over 90% on the intra-rater reliability test. At the conclusion, the development team agreed that the tool was finalized, and each coder could independently code stipulations using the NCI StART tool.

Results: Using StART to Measure CPC CIRB Regulatory Adherence

Following the NCI StART testing process, which ended with the 8th coding round noted above, NCI StART was finalized. The final version of the NCI StART tool was used by four coders to evaluate and categorize the content of stipulations rendered by the CPC CIRB between January 1, 2016 and September 28, 2017. In total, the coders analyzed 1,049 CPC CIRB generated stipulations from 30 protocols and 51 letters reviewed and approved in 2016-2017. If a coder was unsure of the category for a stipulation, the team discussed and determined a final coding assignment.

The objectives of this exercise were: to assess whether the stipulations contained in CPC CIRB determination letters had adhered to the regulatory decision criteria; and to identify any evidence of “overreach,” that is, mandating stipulations that have no reasonable relationship to regulatory requirements, or of “wordsmithing” practices, that is, requiring changes to text based on preference without improving clarity or content.

Use of NCI StART during this exercise demonstrated that the CPC CIRB was using its oversight authority appropriately in accordance with regulations and was not engaged in overreach or wordsmithing. In addition, the tool reduced the potential for subjective uncertainty by including operational definitions and examples for each category and subcategory to facilitate ease of use and increased reliability in correctly assigning the wide range of content found in the stipulations. For example, NCI StART classified “understandable language” in the consent form as an IRB approval criteria stipulation that encompasses the use of technical language or medical jargon, reading level, and translated informed consent documents.

Eighty percent of stipulations were categorized as directly related to IRB approval criteria and 20% were not related to IRB approval criteria. While 20% of the stipulations were not related to one or more IRB approval criteria, these stipulations nevertheless needed to be addressed prior to granting IRB approval because the stipulations had a bearing on the IRB’s overall understanding of the research being reviewed. For example, at times the CPC CIRB issued stipulations relating to inconsistencies between or among documents, e.g., reconcile discrepancy related to the overall study duration length that participants are being asked to participate (12 weeks vs. 24 weeks) and ensure that this information is consistent across all study documents. Although this stipulation is not directly related to IRB approval criteria, the stipulation is appropriate in that the stipulation needed to be addressed in order for the IRB to have a complete understanding of the research being proposed.

Figure 1 shows the distribution of the 1,049 stipulations across the 11 high level codes. Overall, stipulations were most frequently assigned codes from the high-level coding categories of Informed Consent (n=390) and Risks Minimized (n=303). A review of all “compound stipulations” revealed that only IRB approval criteria categories were involved, therefore, this category was included as the ninth IRB approval criteria. The most frequently occurring non-IRB approval categories were: inconsistencies (n=105) and editorial (n=96), accounting for 10% and 9% of all stipulations, respectively. The non-categorizable category, an indicator of scope creep or wordsmithing, accounted for 0.9% of all stipulations.

Stipulations overall occurred most frequently in the informed consent document (48.9%), followed by the protocol document (44.5%), and CIRB application (6.6%). Table 3 shows the distribution of the stipulations to the 22 categories and subcategories; the frequency of each NCI StART category and sub-category; and the location of the stipulation. Within the consent documents, the most frequently coded sub-category was “ICD Elements” (64%) representing stipulations that captured missing consent elements or required improvements to specific elements. Within the protocol document, stipulations were most frequently assigned the codes of research design (32%) and research procedures/instruments (29%). Sixty-nine stipulations (6.6% of stipulations overall) were associated with the CIRB Application of which “inconsistencies” was most frequently assigned category (n=24; 35%). Depending on the nature of the stipulation associated with the CIRB Application, the PI could be required to revise responses provided in the CIRB Application.

Discussion

The NCI StART tool was successful in assessing whether a relatively new federal single IRB (sIRB)¹ was issuing stipulations that adhered to regulatory decision criteria, and whether and to what extent they engaged in overreach or wordsmithing. Less than 1% of all stipulations could not be assigned to any regulatory criteria category or the three defined non-regulatory categories: editorial; inconsistencies; and compound stipulations.

As expected, and consistent with the findings of others (Adams et al., 2014; Kent, 1999; Lidz et al., 2012; Sansone et al., 2004; Tsoka-Gwegweni & Wassenaar, 2014; van Lent, Rongen, & Out, 2014), most CPC CIRB stipulations occurred in the informed consent document (48.9%). Among the seven available StART codes for consent stipulations, most of the consent stipulations (63%) were assigned the StART code of “Informed Consent Document Elements.” This code captures entire missing informed consent elements (e.g., no alternatives included in the informed consent document) or parts of missing elements (e.g., some of the known risks noted in the protocol are missing from the consent form), or improvements to the informed consent elements (e.g., clearly delineate research procedures from standard of care procedures). Of note, other authors report a greater prevalence of stipulations in the informed consent. The presence of fewer consent stipulations may be because NCI requires PIs to use the NCI Informed Consent Template which contains specific instructions and limits the types of changes that can be made to standard language. However, despite this requirement, 6% of the informed consent stipulations were the result of non-adherence to the template.

Although most of the stipulations were in the consent form, a significant number (44.5%) were in the protocol document. The NCI StART codes assigned most frequently to stipulations located in the protocol document were pertaining to risk minimization. Specifically, “Research Design” was assigned to 151 of 467 protocol stipulations and “Research Procedures/Instruments” assigned to 135 of 467 protocol stipulations. These stipulations addressed whether there are issues with the overall research plan, including

¹“Single IRB” is defined here as the selected IRB of record that conducts the ethical review for participating sites of the multi-site study.

research design and methodology, that could expose subjects to unnecessary risks as well as whether risks associated with individual research procedures/instruments are adequately identified, evaluated, described and minimized. Kent, Angell, Sansone, and van Lent also described the frequency and nature of stipulations related to the scientific rationale or the study design.

Two categories of stipulations, while unrelated to IRB approval criteria, are worthy of closer examination: editorial stipulations and inconsistencies. Some may argue that since these stipulations are not related to IRB approval criteria and do not impact the regulatory quality of the research, they should not be addressed by an IRB at the time of review; and further, that these are the very kinds of stipulations that result in investigator burden without improving the protection of research participants. However, it is important to note that the mandate for use of sIRBs for multi-institution research poses greater consequences to both the sIRBs as well as the participating research sites when seemingly inconsequential errors are not addressed at initial IRB review. When an informed consent form with typographical errors or a protocol with inconsistencies approved by a sIRB and subsequently distributed to potentially hundreds of sites in a multi-site study, these errors will be replicated across all institutions utilizing the sIRB. From 15 years of NCI CIRB experience (Massett et al., 2018), the process to correct errors in protocols or consents that were not caught prior to sIRB approval is timeconsuming costly or both for the sIRBs and the participating research sites.

sIRB Implications

While the issue of IRB/REC adherence to regulatory approval criteria is a concern for all IRBs/RECs, this issue assumes a greater significance with the recent implementation of the NIH policy on the use of a sIRB for multi-site research (NIH, 2018) and the sIRB requirements in the Revised Common Rule (HHS, 2017). In this new environment, stipulations issued by an sIRB affect every site participating in a multi-site study, as compared to one or a few sites impacted by stipulations rendered by IRBs/RECs having oversight over single or few sites. Given the broad impact of sIRB stipulations, sIRBs need to be more focused on issuing stipulations that are rooted in the regulatory approval criteria as well as editorial and inconsistencies stipulations that will have a downstream impact on the multiple institutions that would be asked to correct editorial or inconsistencies errors/issues that are not identified until after sIRB approval has been granted and study activation has commenced. There is lingering resistance to adopting central or single review for multisite trials based in part on uncertainty about the quality of reviews by other IRBs (Grady, 2015). Results from NCI StART may serve as a demonstration of quality review to alleviate concerns of institutions that are considering relying on a sIRB.

Best Practices

In addition to the intended use of NCI StART as an assessment tool to determine IRB adherence to regulatory approval criteria, the NCI StART tool – either as presented in this article or as modified to address institutional specific interpretations, requirements, etc., - could be utilized by the IRB/REC community in several ways.

First, NCI StART (or a modified version of NCI StART) can be used by investigators when preparing research studies for IRB/REC review. With NCI StART (or a modified version of NCI StART) as a checklist, investigators can pre-empt stipulations by “thinking like the IRB/REC.” Second, NCI StART (or a modified version of NCI StART) can assist IRB/REC staff who are responsible for writing the IRB/REC stipulations contained in determination letters sent to the study PI. Specifically, this prospective use of NCI StART (or a modified version of NCI StART) could provide IRB/REC staff with the context and rationale to frame the stipulations, so it is clear to all parties why the concern is being raised. Moreover, such use of NCI StART (or a modified version of NCI StART) could result in stipulations that are written clearly with no ambiguity that are easily understood by the study team and result in satisfactory responses the first time around to ensure quicker approval of reviewed research.

Third, NCI StART (or a modified version of NCI StART) may benefit IRB/REC members when reviewing research studies. IRB/REC members could use NCI StART (or a modified version of NCI StART) as a quick reference guide when performing their IRB/REC review to ensure that the reviewers consider all applicable approval criteria and that their proposed stipulations align with approval criteria.

It is important to recognize the limitations of this work. Although NCI StART can be used by any investigator conducting research governed by the Common Rule/FDA regulations as well as any IRB/REC staff person/IRB/REC member responsible for reviewing and approving research in accordance with the Common Rule/FDA regulations, NCI StART was designed for use in the NCI’s CPC CIRB. As a result, operational definitions and examples cited in NCI StART may need to be tailored prior to use by another institution. Thus, *if* an investigator/IRB/REC staff person/IRB/REC member would need to revise the Tool prior to use to address institutional specific interpretations, requirements, etc., it is recommended that the institution conduct concordance testing prior to use of the Tool within the institution to ensure consistency in operation. Alternatively, *if* an investigator/IRB/REC staff person/IRB/REC member could use the Tool without revision, concordance testing would not be needed.

Moreover, the data and accompanying analysis emanating from the use of NCI StART should not be generalized to other IRBs/RECs for the following reasons. First, this data is specific to one of the NCI CIRBs, a sIRB. Second, data came from studies that underwent scientific review prior to IRB review; a process that is not available to all researchers seeking IRB review and approval. Third, data is generated from studies emanating from a medical specialty IRB, oncology.

Research Agenda

Additional research on changes stipulated by IRBs/RECs to secure approval should focus on how such stipulations improve the protection of human subjects; are related to IRB approval criteria; and whether they avoid overreach and wordsmithing. Further review and refinement of existing tools along with systematic testing in a variety of IRB/REC settings and studies may provide the additional empiric evidence upon which to build future improvements.

Educational Implications

For sites with newly established IRBs/RECs, the Tool could be used for internal education. For instance, NCI StART could be used as an educational guide by researchers when developing research protocols to ensure that all applicable FDA/Common Rule IRB approval criteria are addressed, thereby potentially resulting in fewer stipulations being issued at the time of IRB/REC review. Moreover, NCI StART could be used as an educational tool to guide both IRB/REC staff and members when issuing IRB stipulations by ensuring that stipulations align with applicable FDA/Common Rule approval criteria. Lastly, NCI StART could be utilized as an educational tool by quality assurance staff to help assess whether research is being designed, approved and conducted in accordance with FDA regulations/the Common Rule.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Sources of support: This article has been funded in whole or in part with federal funds from the National Cancer Institute under Contract Numbers HHSN261200800001E and HHSN261201400023C.

References

- Abbott L, & Grady C (2011). A systematic review of the empirical literature evaluating IRBs: what we know and what we still need to learn. *Journal of Empirical Research on Human Research Ethics*, 6(1), 3–19. doi:10.1525/jer.2011.6.1.3
- Adams P, Kaewkungwal J, Limphattharacharoen C, Prakobtham S, Pengsaa K, & Khusmith S (2014). Is your ethics committee efficient? Using “IRB metrics” as a self-assessment tool for continuous improvement at the faculty of tropical medicine, Mahidol University, Thailand. *PLoS One*, 9(11), e113356. doi:10.1371/journal.pone.0113356 [PubMed: 25406085]
- Angell EL, Bryman A, Ashcroft RE, & Dixon-Woods M (2008). An analysis of decision letters by research ethics committees: the ethics/scientific quality boundary examined. *Quality and Safety in Health Care*, 17(2), 131–136. doi:10.1136/qshc.2007.022756 [PubMed: 18385408]
- Blackwood RA, Maio RF, Mrdjenovich AJ, VandenBosch TM, Gordon PS, Shipman EL, & Hamilton TA (2015). Analysis of the nature of IRB contingencies required for informed consent document approval. *Accountability in Research*, 22(4), 237–245. doi:10.1080/08989621.2014.956866 [PubMed: 25897768]
- Clapp JT, Gleason KA, & Joffe S (2017). Justification and authority in institutional review board decision letters. *Social Science & Medicine*, 194, 25–33. doi:10.1016/j.socscimed.2017.10.013 [PubMed: 29059597]
- Grady C (2015). Institutional review boards: purpose and challenges. *Chest*, 148(5), 1148–1155. doi: 10.1378/chest.15-0706 [PubMed: 26042632]
- Gunsalus CK, Bruner EM, Burbules NC, Dash L, Finkin M, Goldberg JP, ... Pratt MG (2006). Mission creep in the IRB world. *Science*, 312(5779), 1441–1441. doi:10.1126/science.1121479 [PubMed: 16763114]
- Kent G (1999). Responses by four local research ethics committees to submitted proposals. *Journal of Medical Ethics*, 25(3), 274–277. [PubMed: 10390686]
- Krippendorff K (1970). Estimating the reliability, systematic error and random error of interval data. *Educational and Psychological Measurement*, 30(1), 61–70. doi:10.1177/001316447003000105

- Lidz CW, Appelbaum PS, Arnold R, Candilis P, Gardner W, Myers S, & Simon L (2012). How closely do institutional review boards follow the Common Rule? *Academic Medicine*, 87(7), 969–974. doi:10.1097/ACM.0b013e3182575e2e [PubMed: 22622205]
- Masset HA, Hampp SL, Goldberg JL, Mooney M, Parreco LK, Minasian L, ... Abrams JS (2018). Meeting the challenge: The National Cancer Institute's Central Institutional Review Board for multi-site research. *Journal of Clinical Oncology*, 36(8), 819–824. doi:10.1200/jco.2017.76.9836 [PubMed: 29384720]
- Nicholls SG, Hayes TP, Brehaut JC, McDonald M, Weijer C, Saginur R, & Fergusson D (2015). A scoping review of empirical research relating to quality and effectiveness of research ethics review. *PLoS One*, 10(7). doi:ARTN e0133639 10.1371/journal.pone.0133639
- National Cancer Institute. (2018). Division of Cancer Prevention scientific scope. Retrieved from <https://prevention.cancer.gov/about-dcp/scientific-scope>
- National Institutes of Health. (2018). Final NIH policy on the use of a single institutional review board for multi-site research. Retrieved from <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-16-094.html>
- Olsen DP, & Mahrenholz D (2000). IRB-identified ethical issues in nursing research. *Journal of Professional Nursing*, 16(3), 140–148. doi:10.1053/pn.2000.5918 [PubMed: 10860312]
- Sansone RA, McDonald S, Hanley P, Sellbom M, & Gaither GA (2004). The stipulations of one institutional review board: a five year review. *Journal of Medical Ethics*, 30(3), 308–310. doi: 10.1136/jme.2002.002105 [PubMed: 15173369]
- Tsoka-Gwegweni JM, & Wassenaar DR (2014). Using the Emanuel et al. framework to assess ethical issues raised by a biomedical research ethics committee in South Africa. *Journal of Empirical Research on Human Research Ethics*, 9(5), 36–45. doi:10.1177/1556264614553172 [PubMed: 25747689]
- U.S. Department of Health and Human Services. (2017). Federal policy for the protection of human subjects. Retrieved from <https://www.gpo.gov/fdsys/pkg/FR-2017-01-19/pdf/2017-01058.pdf>
- van Lent M, Rongen GA, & Out HJ (2014). Shortcomings of protocols of drug trials in relation to sponsorship as identified by research ethics committees: analysis of comments raised during ethical review. *BMC Medical Ethics*, 15, 83. doi:10.1186/1472-6939-15-83 [PubMed: 25490963]
- White RF (2007). Institutional review board mission creep, The Common Rule, social science, and the nanny state. *The Independent Review*, 11(4).
- Whitney SN (2016). Institutional review boards: A flawed system of risk management. *Research Ethics*, 12(4), 182–200. doi:doi:10.1177/1747016116649993

Figure 1.

Distribution of 1049 stipulations across 11 high-level NCI StART codes

Table 1.

Stipulation Category Explanations, Sub-categories and Examples

CATEGORY EXPLANATION	SUB CATEGORY	EXAMPLE STIPULATION
Category 1: Risks to Subjects Minimized Captures stipulations addressing whether there are issues with the overall research plan, including research design and methodology, that could expose subjects to unnecessary risks as well as whether risks associated with individual research related procedures/instruments are adequately identified, evaluated, described and minimized.	1A. Research Design	The primary objective of this study is to test for the equivalency of the two arms. This should be incorporated into the statement of the primary objective to reflect the study as an equivalency trial.
	1B. Research Procedures/Instruments	Please clarify the following sentence to allow for the optional biopsies at week 9. "Participants may be withdrawn from the study if they are not able to provide blood, urine or tissue samples, or they are not able to take their study medication as instructed."
	1C. Diagnostic or treatment procedures being used for research purposes	Please explain why individuals need to undergo a research indicated CT scan when the individuals will be undergoing a standard of care diagnostic CT scan within 5 days of the research indicated CT scan.
Category 2: Risks Reasonable in Relation to Anticipated Benefits Captures stipulations addressing whether the protocol adequately identifies, evaluates and describes the potential benefits that may be gained by current subjects; whether and how knowledge gained from the research may benefit current subjects, future subjects or society and whether the potential benefits to subjects or society outweigh risks inherent in the research, i.e., whether the risk/benefit ratio is acceptable.	2A. Risk Reasonable in Relation to Anticipated Benefits	In the application and the consent form, the study team indicates that there is no benefit to study participation. Consideration should be given to including as a benefit an examination of tissue biopsies for abnormalities that otherwise would not have been performed.
Category 3: Equitable Selection of Subjects Captures stipulations addressing whether appropriate subjects are being recruited into the research as well as how they are being recruited/enrolled is appropriate.	3A. Inclusion/Exclusion Criteria	Either revise the inclusion/exclusion criteria, or provide further scientific rationale to justify why the following groups are excluded from participating in the study: a) Males b) HIV positive patients c) Patients who received prior treatment.
	3B. Screening/Recruitment/Enrollment	The recruitment plan describes a variety of recruitment methods that will be used to recruit participants – all recruitment materials and plans to use social media for recruitment must be submitted to the IRB for review and approval before implementation.
Category 4: Informed Consent Captures stipulations addressing whether the process of obtaining informed consent is adequate and whether the Consent Form contains all the required elements.	4A. Process	During the consent for the pre-operative clinic visit, it should be recommended that the participant be accompanied by a family member or friend and at each visit prior to surgery.
	4B. Understandable Language	Explain 'dyspepsia' in lay language.
	4C. Exculpatory Language	The statement "I waive any possibility of compensation, including any right to sue, for injuries that I may receive because of participation in this research" is considered exculpatory where a participant forfeits their rights for compensation for injury and the right to sue. This statement must be deleted.
	4D. Coercion or Undue Influence	The sentence, "Studies in women treated with breast conserving surgery have demonstrated equal to improved breast appearance, skin redness and long term arm swelling with short course radiation schedules," could be considered coercive. The sentence could be interpreted as indicating the study treatment is better than the usual approach. Please delete this sentence.

CATEGORY EXPLANATION	SUB CATEGORY	EXAMPLE STIPULATION
	4E. ICD Elements	Include all the procedures which occur as part of participation in the research study, from screening to last follow-up visit/contact.
	4F. Additional Information to be provided to participants	While the CIRB recognizes that the sexual partner would not be considered a research participant at any point of the study, the IRB recommends developing an information sheet or educational material that the participant could give to their sexual partner if there are any safety concerns with drug exposure and sexual intercourse.
	4G. Adherence to a consent form template	In the section “Who can answer my questions about this study” use the template language.
Category 5: Waiver or Alteration of Informed Consent Process or Signed Document Captures stipulations addressing waiver or alteration of the informed consent process or a waiver or alteration of documentation of Informed Consent.	5A. Waiver/Alteration of documentation of IC Process	If you request a waiver consent for prescreening activities provide the IRB with a script or outline of the conversation that will take place to obtain consent.
	5B. Waiver/Alteration of documentation of IC	Before the IRB can approve a waiver of consent for the screening research activity, please explain how the screening activity: a) Involves no more than minimal risk, b) Would not be practical without the waiver c) Does not adversely affect the potential participants’ rights and welfare, and d) Whether participants will be provided with pertinent information about participation in the study as part of the overall recruitment procedures.
Category 6: Data Monitoring to Ensure Subject Safety, When Appropriate Captures stipulations addressing the monitoring of research data.	6A. Data Safety Monitoring Plan/Board	Include probability-based safety boundaries or monitoring rules or guidelines in the event of an unexpectedly large number of very serious Adverse Events (AE’s) (e.g. deaths or hospitalizations or selected AE’s that participants with this disease could experience).
Category 7: Privacy and Confidentiality, When Appropriate Captures stipulations addressing whether there are adequate privacy and confidentiality protections included, adequately described and justified.	7A. Privacy and Confidentiality	The IRB requests that a NIH Certificate of Confidentiality (CoC) be obtained for this study to help further assure confidentiality and privacy to the study participants.
Category 8: Vulnerable Subjects Captures stipulations addressing issues that are germane to vulnerable population, e.g., educationally disadvantaged individuals and whether those individuals would be able to read the ICD.	8A. Vulnerable Subjects	Section 2.4.1(d) notes that persons with impaired decision-making capacity are eligible to participate in the study; however, the rest of this section was not completed. Complete the rest of checkboxes in this question including providing details of the additional safeguards for the protection of this population.
Category 9: Editorial Captures stipulations related to correcting spelling, grammar, typing or formatting.	9A. Editorial	In the Additional Studies Section, under “What is involved”, item #6, change “the” to “they” in the second sentence.
Category 10: Inconsistencies Captures stipulations addressing inconsistencies between documents, e.g., consent and protocol, or within documents, e.g., one section of the document says one thing and another section says something different, and requires correction.	10A. Inconsistencies	Reconcile discrepancy related to the overall study duration length that participants must be in the study (12 weeks vs 24 weeks post-surgery) and ensure that this information is consistent throughout all study documents.
Category 11: Non Categorizable Captures stipulations that cannot be categorized under any of the preceding categories, e.g., IRB overreach, IRB wordsmithing, etc.	11A. Cannot assign to category	Change “subjects” to “participants” throughout the consent form when referring to the individuals who will be participating in the research.
Category 12: Compound Stipulations Captures stipulations covering more	12A. Compound Stipulations	Please include a formal plan to have participants restrict their use of turmeric and anticoagulants while on the study. Also, please revise the following

CATEGORY EXPLANATION	SUB CATEGORY	EXAMPLE STIPULATION
than 1 IRB approval criteria coded category.		statement because participants receiving anticoagulants are not eligible for this study: "Therefore it may be beneficial to restrict those subjects on anticoagulant therapy from using the supplementation."

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2.

Results of concordance and reliability testing for NCI StART

Testing Round	Tool Version	# of protocols/ Board	# of stips coded	# of coders	% concordance	Krippendorff Alpha	Date
1	1.0	10 CPC	299	2	67%	---	2/6/16
2	2.0	10 CPC	299	2	49%	---	3/31/16
3	3.2	3 CPC	80	2	70%	---	4/15/16
4	3.2	3 CPC	80	4	21%	---	5/3/16
5	4.1	3 CPC	62	4	61%	0.41	6/24/16
6	5.0	3 CPC	72	4	65%	0.739	7/13/16
7a	6.1	5 (Peds)	48	4	91%	0.563	8/4/16
7b	6.1	3 (LPE)	58	4	84%	0.587	8/4/16
8	7.1	2 (EPE)	76	4	92%	0.869	8/11/16

Table 3.

NCI StART coding results (N=1,049)

Code	NCI StART Category	Protocol	Consent	Application
		N=467 n (%)	N=513 n (%)	N=69 n (%)
1A	Research Design	151 (32)	(N/A)	3 (4)
1B	Research Procedures/Instrument	135 (29)	(N/A)	13 (19)
1C	Diagnostic or Tx Procedures being used for research purposes	1 (0.2)	(N/A)	0 (0)
2A	Risks Reasonable in relation to anticipated benefits	2 (0.4)	(N/A)	1 (1.4)
3A	Inclusion/Exclusion Criteria	61 (13)	(N/A)	6 (8.7)
3B	Screening/Recruitment/Enrollment	21 (4)	(N/A)	2 (2.8)
4A	IC: Process	1 (0.2)	2 (0.4)	0 (0)
4B	IC: Understandable Language	0 (0)	25 (4.9)	0 (0)
4C	IC: Exculpatory Language	0 (0)	1 (0.2)	0 (0)
4D	IC: Coercion and Undue Influence	0 (0)	2 (0.4)	0 (0)
4E	IC: ICD Elements	1 (0.2)	326 (63)	1 (1.4)
4F	IC: Additional information to be provided to participants	0 (0)	2 (0.4)	0 (0)
4G	IC: Adherence to IC Template	0 (0)	29 (6)	0 (0)
5A	Waiver/Alteration of IC Process	1 (0.2)	1 (0.2)	0 (0)
5B	Waiver/Alteration of documentation of IC	0 (0)	1 (0.2)	0 (0)
6A	Data Safety Monitoring Plan/Board	10 (2)	0 (0)	0 (0)
7A	Privacy and Confidentiality	4 (0.9)	6 (1.2)	2 (2.8)
8A	Vulnerable Subjects	0 (0)	0 (0)	3 (4)
9A	Editorial	21 (4)	63 (12)	12 (17)
10A	Inconsistencies	40 (8)	41 (8)	24 (35)
11A	Cannot assign to category	5 (1)	3 (0.5)	2 (2.8)
12A	Compound Stipulations	13 (3)	11 (2)	0 (0)