



Published in final edited form as:

*Clin Trials*. 2014 August ; 11(4): 467–472. doi:10.1177/1740774514538706.

## Screen Failure Data in Clinical Trials: Are Screening Logs Worth It?

Jordan J. Elm, PhD<sup>1</sup>, Yuko Palesch, PhD<sup>1</sup>, J. Donald Easton, MD<sup>2</sup>, Anne Lindblad, PhD<sup>5</sup>, William Barsan, MD<sup>3</sup>, Robert Silbergleit, MD<sup>3</sup>, Robin Conwit, MD<sup>4</sup>, Catherine Dillon, CCRP<sup>1</sup>, Mary Farrant, MBA, RN<sup>2</sup>, Holly Battenhouse, MS<sup>1</sup>, Aaron Perlmutter, MS<sup>1</sup>, and S. Claiborne Johnston, MD, PhD<sup>2</sup>

<sup>1</sup>Department of Public Health Sciences, Medical University of South Carolina

<sup>2</sup>Departments of Neurology and Epidemiology and Biostatistics, University of California, San Francisco

<sup>3</sup>University of Michigan

<sup>4</sup>National Institute of Neurological Disorders

<sup>5</sup> The EMMES Corporation

### Abstract

**Background**—Clinical trials frequently spend considerable effort to collect data on patients who were assessed for eligibility but not enrolled. The Consolidated Standards of Reporting Trials (CONSORT) guidelines’ recommended flow diagram for randomized clinical trials reinforces the belief that the collection of screening data is a necessary and worthwhile endeavor. The rationale for collecting screening data includes scientific, trial management, and ethno-socio-cultural reasons.

**Purpose**—We posit that the cost of collecting screening data is not justified, in part due to inability to centrally monitor and verify the screening data in the same manner as other clinical trial data.

**Methods**—To illustrate the effort and site-to-site variability, we analyzed the screening data from a multi-center, randomized clinical trial of patients with transient ischemic attack or minor ischemic stroke (POINT).

**Results**—Data were collected on over 27,000 patients screened across 172 enrolling sites, 95% of whom were not enrolled. Although the rate of return of screen failure logs was high overall (95%), there were a considerable number of logs that were returned with “no data to report” (23%), often due to administrative reasons rather than no patients screened.

**Conclusions**—In spite of attempts to standardize the collection of screening data, due to differences in site processes, multi-center clinical trials face challenges in collecting those data

---

**Correspondence to** Jordan J. Elm, PhD, Department of Public Health Sciences, Medical University of South Carolina, 135 Cannon St., Suite 303, PO Box 250835, Charleston, SC 29425, Phone (843) 876-1605, Fax (843) 876-1126, elmj@musc.edu.

The authors declare that there is no conflict of interest.

completely and uniformly. The efforts required to centrally collect high-quality data on an extensive number of screened patients may outweigh the scientific value of the data. Moreover, the lack of a standardized definition of “screened” and the challenges of collecting meaningful characteristics for patients who have not signed consent limits the ability to compare across studies and to assess generalizability and selection bias as intended.

### Keywords

Consolidated Standards of Reporting Trials guidelines; Screening logs; Screen Failure; TIA; minor ischemic stroke

---

### Introduction

The first figure in many randomized clinical trial primary manuscripts is a flow diagram of patient enrollment and follow-up. This diagram is recommended by the Consolidated Standards of Reporting Trials (CONSORT) Statement.<sup>1</sup> The standard format includes in the top box the number of patients “assessed for eligibility”, followed by the subset who were excluded, and the subset who were randomized.<sup>1, 2</sup> The inclusion of the number of patients screened in the flow diagram in the 1996 CONSORT statement<sup>2</sup> has been criticized as being of little meaningful scientific value.<sup>3</sup> Although it remained in the revised 2010 CONSORT statement flow diagram, the authors acknowledge that this count may be unknown or may not be as valuable as other counts.<sup>4</sup>

The operational definition of *screened patients* and the approach to the collection of screening data vary widely across studies and clinical sites. Currently, there is no standardized way to define who was assessed for eligibility, or *screened*. *Screened patients* could broadly be defined as patients with the disease who present at the site(s) during the recruitment time interval, including those who were not formally assessed for eligibility. In contrast, *screened patients* could be defined narrowly as those who sign informed consent. If defined too broadly, the task of reporting screen failures is resource exhaustive. If defined too narrowly, the scientific merit of the screening data collected is lost. If inconsistently collected, the data are not interpretable. A review of stated purposes for collecting screen failure data can help evaluate whether these data, as commonly reported, actually add value within the clinical trial enterprise. The purpose for collecting screen failure data may include the following: 1. scientific reasons, 2. trial management, and 3. ethical and socio-cultural considerations.

### Scientific Reasons

The main purpose of the CONSORT guidelines is to ensure that the scientific validity of the trial can be demonstrated in the published report. The CONSORT group's explanation for reporting screening counts is as follows:

“If available, the number of people assessed for eligibility should also be reported. Although this number is relevant to external validity only and is arguably less important than the other counts, it is a useful indicator of whether trial participants were likely to be representative of

all eligible participants” and then subsequently “these counts indicate whether trial participants were likely to be representative of all patients seen...”<sup>4</sup>

The authors’ explanations conflate the goals of showing whether enrolled population is a large or small subset of those eligible for enrollment, or a large or small subset of those patients seen with the medical condition of interest. Indeed this confusion reflects the lack of a consistent and useful definition for entry into screening logs. The general implication of this statement is that by collecting screening data it will be possible to demonstrate the generalizability of the findings and lack of selection bias in the study subjects. This premise is, however, a relatively crude assessment since the counts themselves tell us nothing about actual specific clinical characteristics of patients with the disease of interest not included in the study. Assessments of those eligible are further limited as patients who decline may or may not have been eligible, and all aspects of eligibility may not be assessed for every patient screened.

For trial results to be generalizable, patients enrolled should be a representative sample of the target patient population at large. Overly narrow eligibility criteria, therefore, limit external validity, but it is important to distinguish between exclusion criteria that are expected to affect clinical generalizability from those that are merely pragmatic. Some eligibility criteria are for practical reasons (e.g. currently receiving another experimental treatment or not expected to be available for follow-up) which would not limit the generalizability, per se. Screening that does not accurately identify specific reasons for exclusion may, therefore, also be misleading.

It is possible that patients with a uniformly better, or uniformly worse, prognosis are given priority for enrollment based on the investigator's perception of whether the study may be “right” for the patient. In most trials, patients self-select to participate. Some have argued that screening data are useful to detect selection bias as demonstrated for two multicenter studies of traumatic brain injury. However, the differences observed were largely expected due to the trials’ eligibility criteria or were sampling biases indirectly induced by site practices outside of the control of the investigator.<sup>5</sup> While it may be meaningful to identify unexpected differences between enrolled and non-enrolled patients, it may not be possible to determine the reasons for these differences or ascertain whether they will result in biased inference.

## Trial management

For multi-center trials, the collection of screening data is a trial management tool. Frequently, clinical trials face challenges in recruitment and, increasingly, research funding is directly tied to meeting the expected recruitment timelines. The centralized collection of screening data helps ensure that all sites are actively screening for the study and are correctly ruling out potential subjects. Screening data collected from high enrolling sites may provide model examples for other sites. Sites with low recruitment rates and low numbers of patients screened may have limited hours of availability for enrollment, may not adequately be identifying potential subjects, or may lack the required patient population for study participation. For trials with difficulty recruiting across most sites, an analysis of the primary

reason(s) for exclusion may identify key eligibility criteria that are problematic. The reasons for which eligible patients decline participation may help to understand barriers to study participation or to compare methods of obtaining informed consent. For example there may be barriers, such as protocol intensity, travel or time, which are common across all patients or specific for particular subgroups, e.g. amongst minority or elder populations, which can be remedied by providing travel reimbursements, allowing remote assessments, or streamlining data collection.

## Ethical and socio-cultural

Screening data can help to determine whether the Belmont principle of justice is being applied; all potential research participants are equally sharing the costs and benefits of research participation.<sup>6</sup> Sponsors, regulators, and human-subjects review boards encourage the rates of recruitment to be as expected across gender, racial and ethnic groups. Ethics committees are charged with ensuring that trials do not unfairly exploit or exclude certain subgroups of patients. In the absence of screening data it may be difficult for investigators to know that study processes are enrolling subjects in a disproportionate manner. Screening data may be useful for investigators that are having difficulty recruiting a representative demographic sample to identify specific reasons that members of different groups may decline participation. However, this assumes that the true reasons for declining can be ascertained, which may not be possible if failure to build a personal connection is the reason for the lack of willingness to participate.

## Approaches and Challenges to Collecting Screening Data

The process of documenting screening activities may be a useful exercise to help sites to understand or improve the quality of their screening efforts. There is value in assessing screening data at the site level to identify if potentially eligible subjects are being missed, if the outcomes of informed consent processes for some investigators or coordinators are anomalous, or if they are recruiting aberrantly high or low rates of certain demographic groups. Ideally screening logs simply reinforce and centrally collect information that is already recorded as part of best site management practices. However, once a site has confirmed that it is not missing patients based on its local processes, the screening logs are unlikely to add value when weighed against the resources required to collect the data. A high performing quality assurance or improvement program at sites may involve screening at the outset of recruitment and then at sampling intervals, or for cause, as the trial progresses.

In practice, there is substantial between-site variability in how screening data are collected and for whom. In an effort to standardize the collection of data across studies, the National Institute of Neurological Diseases and Stroke developed Common Data Elements (CDEs). The CDE screening log instructions are to include “all individuals who entered pre-screening or screening”.<sup>7</sup> This definition needs to be further defined for a specific protocol. Moreover, a more useful standard for inclusion in screening logs would be presentation with the disease of interest, i.e., the population that is intended to be screened for the study, rather than the subjective, circular process definition of having been screened.

The accuracy of the reported reasons for screen failures must be called into question. To simplify data collection, the site is often limited to picking the primary reason for ineligibility or declining consent, although there could be more than one. Permitting the documentation of multiple reasons for ineligibility will not fully remedy the problem of over-reporting early eligibility indicators, because screening efforts typically stop once ineligibility has been determined. Certain eligibility criteria take longer to ascertain than others, and this results in a bias against the reporting of reasons that are more resource intensive to assess.

Unlike study research data, screening data is usually not centrally monitored nor verified against source documents. Ethical and regulatory restrictions do not allow collection of the identifiers needed for this kind of monitoring, nor is it clear that the imprecise questions of external validity asked of these data necessitate the same level of scrutiny as do the data essential to the primary objectives of the study. Attitudes of investigators towards screening data also vary. If the purpose of collecting screening data is perceived to be only for trial management, then the coordinating center may not care whether screening data are being reported, or reported accurately, by high enrolling sites.

The Health Insurance Portability and Accountability Act (HIPAA) rules restrict the collection of individually identifiable health information without consent. This restriction hampers the ability to collect meaningful information about patients who were screened.<sup>9</sup> Site-to-site variability is further increased by each local Institutional Review Board's (IRB's) interpretation of the acceptability of reporting characteristics of patients who have not signed consent. Strict interpretation can severely limit the ability to demonstrate the scientific validity of the findings (i.e., generalizability without selection bias). Others have described the challenges of collecting clinical characteristics for non-consented patients.<sup>8</sup> They note that among US sites, IRBs are more likely to object to the collection of clinical characteristics for non-consented patients.<sup>5,8</sup> Multi-center clinical trials either have to limit the type of data collected on screen failure logs to be acceptable to the most restrictive IRBs, or accept the fact that some sites will not be allowed to submit the screening data due to local IRB requirements.

### **An Example of Site-to- Site Variability in the Collection of Screening Data**

To illustrate the challenges of collecting screening data, the screening data from the Platelet-Oriented Inhibition in New Transient Ischemic Attack and Minor Ischemic Stroke (POINT) Trial was analyzed (ClinicalTrials.gov Identifier:NCT00991029). The POINT trial is a double-blind, randomized, multicenter study of clopidogrel in patients with transient ischemic attack or minor ischemic stroke. The trial is currently enrolling, and data are reported here as of January 24, 2013 at which time 1,285 patients were enrolled. Sites were asked to report “patients who were actively screened (in person or via telephone) for the POINT study by your study team but not randomized”. A total of 27,292 patients were screened among 172 sites, and less than 5% of these were enrolled (Table 1). The sites were grouped into tertiles based on the percent of their target recruitment that was achieved. The target recruitment depended on the length of time the site had been activated to enroll patients. The sites in the top enrollment tertile enrolled the majority of the patients (72%),

and 6.3% of all the patients screened by top enrolling sites were actually enrolled. Lower performing sites, in the middle and bottom tertile of enrollment, recruited a smaller percentage of screened patients (3.4% and 1.6%, respectively). The high enrolling sites screened on average about three times as many patients as the low enrollers (12 versus 4 patients screened per month).

The completeness of screening data in the POINT trial is shown in Table 2. Overall, 95% of the expected monthly screening logs were returned; however, 23% of the time there were no screen failures to report for the month. When queried these sites claimed no activity for the month. Null values in this scenario have a different meaning than truly no potential patients were available for screening. For example, the minimum number of patients screened per month is zero when averaging across monthly logs, suggesting that for at least one site, no one failed to meet enrollment criteria, which is not possible (Table 1). High enrolling sites were less likely to return the monthly screening log with no screen failures to report, (Table 2). This association between lower enrolling sites and lower return of screen failure logs has also been noted in other settings.<sup>5</sup>

Given the inclusion/exclusion criteria, it was expected that the vast majority of patients with stroke/TIA would not be eligible, and this is confirmed by the high proportion of patients who were screened and not enrolled. These data do suggest that the POINT trial patients represent a low percentage of the overall minor stroke/TIA population. Any further inference is limited because considerable portions of the screening data are missing, and many low enrolling sites are not reporting screening data as expected. Even with careful training, these data likely will still be incomplete.

## Recommendations

The purposes for collecting screen failure data are multifaceted, but the considerable challenges, inefficiencies, and limitations to the way screening logs are commonly used today make it difficult to attain any of these goals. Consideration of improvements in standard practice, or eliminating screening logs entirely and finding alternatives are warranted.

A combination of best practices and innovations in the way screening logs are collected may improve their utility. If screening logs are used in a trial, their use should be mandatory regardless of whether intended to address issues of external validity, optimize trial management, or identify demographic imbalances. Entry criteria for screening logs must be defined for each trial in a consistent manner appropriate to the patient population to whom external validity is to be demonstrated. Ideally the to-be-screened population would be identifiable through existing administrative datasets such as admission logs or billing codes. Sampling strategies rather than absolute counts could be used to improve efficiency and may allow more slightly granularity for the data collected. Screening logs should remain restricted in content, but should document as many eligibility criteria as practical.

Even when definitions and processes for collecting screening data are clear, the challenges of centrally collecting information on non-consented patients limit the ability to ensure high

quality data. Simply reporting the total number assessed for eligibility, the number ineligible, the number declined, and the number excluded, as recommended by CONSORT, does not clearly demonstrate external validity. Alternatives to screening logs may be more effective than trying to fix them. Selection bias might be better assessed by baseline comparisons with other similar clinical trials. Moreover, generalizability might be better assessed by comparing baseline characteristics of trial participants to the body of existing literature of the disease characteristics/profile. Given the need to decrease the costs of conducting clinical trials, the cost versus benefit of each data collection element should be carefully assessed.

## Acknowledgements

The authors thank the patients and families who participate in the POINT study and the POINT Investigators for the data collected. This study was sponsored by grants from the National Institute of Health/National Institute of Neurological Disorders and Stroke [grant numbers U01NS062835 and U01NS059041].

## References

1. Schulz KF, Altman DG, Maher D, the CONSORT Group. CONSORT. 2010 Statement: Updated Guidelines for Reporting Parallel Group Randomized Trials. *Ann Intern Med.* 2010; 152:726–732. [PubMed: 20335313]
2. Begg C, Cho M, Eastwood S, et al. Improving the quality of reporting of randomized controlled trials: the CONSORT statement. *JAMA.* 1996; 276:637–639. [PubMed: 8773637]
3. Meinert CL. Beyond CONSORT: Need for Improved Reporting Standards for Clinical Trials. *JAMA.* 1998; 279:1487–1489. [PubMed: 9600487]
4. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, Elbourne D, Egger M, Altman DG, the CONSORT Group. CONSORT. Explanation and Elaboration: updated guidelines for reporting parallel group randomised trial. *BMJ* 2010. 2010; 340:c869. doi: 10.1136/bmj.c869.
5. Sliker FJA, Kompanje EJO, Murray GD, et al. Importance of screening logs in clinical trials for severe traumatic brain injury. *Neurosurgery.* 2008; 62:1321–9. [PubMed: 18824999]
6. National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research. <http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.html> (1979, accessed 14 May 2013)
7. Grinnon ST, Miller K, Marler JR, Lu Y, Stout A, Odenkirchen J, Kunitz S. National Institute of Neurological Disorders and Stroke Common Data Element Project - approach and methods. *Clin Trials.* 2012; 9:322–9. [PubMed: 22371630]
8. Kompanje EJO, Maas AIR. Is the Glasgow Coma Scale score protected health information? The effect of new United States regulations (HIPAA) on completion of screening logs in emergency research trials. *Intensive Care Med.* 2006; 32:313–4. [PubMed: 16468073]
9. Maas AIR, Kompanje EJP, Sliker FJA, Stocchetti N. Differences in completion of screening logs between Europe and the United States in an emergency phase III trial resulting from HIPAA requirements. *Ann Surg.* 2005; 241:382–383. [PubMed: 15650653]



**Table 1**

Screened versus Enrolled Patients by Site Performance (POINT Trial)

Sites	Number of Sites	Number of Patients Screened	Average number of patients screened* per month (min-max)	Number of Patients Enrolled (% of Total Enrolled)	% of patients screened but not enrolled (% of total screened)
Total	172	27,292	8 (0-98)	1,285 (100%)	95.2%
Low Enrollers (bottom tertile 0-28% of target)	57	3291	4 (0-50)	52 (4%)	98.4%
Medium Enrollers (mid tertile 28%-74% of target)	58	9093	8(0-66)	305 (24%)	96.6%
High Enrollers (top tertile 75%+ of target)	57	14908	12 (0-98)	928 (72%)	93.7%

NOTE: Numbers are as of Jan 24, 2013. Enrollment is ongoing. Table excludes 5 sites released to enroll within 2 months of Jan 24, 2013 (the date of the data freeze).

\* screened but not enrolled

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Table 2**

Completeness of Screening Data in the POINT Trial

	N of monthly logs expected*	N (%) of monthly screening logs completed with data	N (%) screening logs completed as "No data to Report"	N (%) logs not completed	N (%) of sites providing no screening data (ever)
Total	3239	2330 (72%)	750 (23%)	159 (5%)	10 (6%)
Low Enrollers (bottom tertile 0-28% of target)	832	447 (54%)	339 (41%)	46 (6%)	8 (14%)
Medium Enrollers (mid tertile 28%-74% of target)	1188	877 (74%)	262 (22%)	49 (4%)	1 (2%)
High Enrollers (top tertile 75%+ of target)	1219	1006 (83%)	149 (12%)	64 (5%)	1 (2%)

\* Expected based on number of months since the site was "released to enroll" up to Dec 2012.