Making Prospective Registration of Observational Research a Reality

Rafael Dal-Ré, John P. Ioannidis, Michael B. Bracken, Patricia A. Buffler, An-Wen Chan, Eduardo L. Franco, Carlo La Vecchia, Elisabete Weiderpass

The vast majority of health-related observational studies are not prospectively registered and the advantages of registration have not been fully appreciated. Nonetheless, international standards require approval of study protocols by an independent ethics committee before the study can begin. We suggest that there is an ethical and scientific imperative to publicly preregister key information from newly approved protocols, which should be required by funders. Ultimately, more complete information may be publicly available by disclosing protocols, analysis plans, data sets, and raw data.

Although observational studies (OSs) of human subjects are considered to be more practical and closer to real-life daily practices than are clinical trials (CTs), the transparency that the scientific community demands for the latter is considered irrelevant for OSs. Recommendations that OSs should be publicly registered, as occurs for CTs, has been debated for years, with supporting, dissenting, and neutral views (1–8). Some investigators conducting epidemiological OS already register their studies: As of 29 January 2014, 29,826 OSs are registered in ClinicalTrials.gov (18.6% among more than 160,000). However, the great majority of OSs are unregistered, despite the fact that OSs represent a large fraction of published health-related research.

SURVEYING THE LANDSCAPE

In order to understand the types of studies that are currently registered and published in the biomedical literature, we searched PubMed for papers that (i) had abstracts; (ii) pertained to humans; (iii) were not classified as reviews or systematic reviews, meta-analyses, case reports, guidelines, editorials, or consensus statements; and (iv) were published in 2011. Of 400,601 such papers, 23,350 (6%) were classified by PubMed as randomized controlled trials” (RCTs), and 377,251 did not have this tag. Using a random sample of 50 of the 377,251 papers, we determined that 36 (72%) were nonrandomized studies; only two listed a registration number, of which one was a single-arm clinical trial and the other a large case-control study. Using a random sample of 0.5% (n = 118) of the 23,350 papers, we found that 89 (75%) were indeed RCTs and 11 (9%) had OS design but used data from RCTs; 22 PubMed records had a listed registration number. These observations suggest that publications of results from OSs far outnumber publications from RCTs, with almost 300,000 versus 20,000 publications of the two study design types per year, respectively. Registration numbers still accompany only ~20% of RCTs, and registration of OSs is distinctly uncommon; one exception may be OSs that use data from RCTs. We found five such cases in our sampling, but the registration record always referred to the original RCT and provided no meaningful information on the study design of the specific OS.

In observational research, it is sometimes difficult to define what constitutes a single study: an analysis, a set of analyses, or a protocol. However, the majority of OS results appearing in peer-reviewed journals are unanticipated; that is, only a small set of authors, funders, and reviewers know about the study’s existence before publication.

PROS AND CONS

There are several postulated benefits in systematically registering all OSs: increasing transparency and credibility, improving the peer-review process and ethical conduct of studies, and ensuring that the totality of evidence is publicly available (1). Moreover, registration of OSs may enhance communication regarding explored, but not published, hypotheses (3), facilitate systematic reviews and research collaborations (8), and reduce redundancy and funding committed to research questions for which adequate studies have already been conducted or are being performed (9), allowing published evidence to be better placed in context.

Registration of CTs started as a means of deterring (and detecting) selective reporting (10). OSs with negative results are believed to be less frequently reported than those with positive findings (6). Indeed, nearly all published OS results report some significant findings, and many report implausibly large effects (11). This
high rate of likely false-positive findings has led policy-makers and the public to regard OS results with skepticism (12). Young and Karr (13) found that none of 52 major claims from selected health-related OSs were validated in RCTs. Swaen et al. (14) showed that OSs without pre-specified hypotheses are more likely to yield published false-positive results than are OSs based on predefined hypotheses. Simply being “hypothesized” does not, of course, itself infer greater validity to a study observation, but hypothesis-based research is more likely than post hoc analysis to involve study populations appropriate for testing the hypothesis, to measure relevant exposures and outcomes in necessary detail, and to collect pertinent covariate information (7). Moreover, hypothesis-based research is more likely to be appropriately powered and thus less likely to produce inflated associations (15). Therefore, some (7, 14), but not all (3, 16), authors speculate that conducting OSs with publicly declared prespecified hypotheses might reduce the rate of false-positive findings. Prospective registration of OSs also may encourage investigators to publish study results irrespective of whether they meet some nominal level of statistical significance because publication is a basic requirement of research ethics (17) and scholarship.

A potential disadvantage of registering highly exploratory, hypothesis-generating research with complex, meandering analyses is the burden of ongoing serial amendments (1, 6, 8) and the resulting hindrance of new idea generation (4, 8), as well as reduction in the analyses of end points not prespecified because they were conceived after the study started (18). However, there is no compelling reason why new concepts should be hindered; they just need to be identified as post hoc observations. Such disclosure allows others to fully understand and openly debate the nature and merit of the analyses. There is no evidence that registering CTs has led to fewer hypotheses being tested or a decline in secondary analysis of trial data. Conversely, there is greater recognition that hypotheses and analyses for testing them need to be specified a priori (19, 20); without such delineation, study results can lead to biased reframing of the hypothesis or cherry-picking among unspecified end points.

Another concern is that competitors may “steal” novel ideas from a registry, complete the analyses, and publish the results first. However, wide adoption of OS registration and transparent sharing of OS data will document primacy of investigators who first registered the idea for a given data set. The act of registering a protocol and study hypotheses can serve as a claim to the conceptual underpinnings of a study.

### THE REGISTRATION CHALLENGE

In contrast to CTs, which require prospective data collection and follow-up of participants, analyses of some OSs can be performed readily in minimal time whenever required data have been collected, perhaps as part of a prior survey or a byproduct of health care activities (for example, administrative or billing databases, disease registries). In these cases, registering a protocol or a full analysis plan may not qualify as prospective. Theoretically, an investigator can mine the available data, notice some provocative results, and build a protocol and analysis plan around the selected results while spuriously claiming that the plan was prospectively conceived. Therefore, for existing OS data sets it is essential to publicly disclose what variables have been collected and are available for analysis. Such disclosure would define, at least in part, the analytical space in which analyses can be readily performed.

Epidemiologists, their professional organizations, and journal editors should endorse OS registration (1) because the reasons supporting registration of CTs outlined in the Ottawa statement (21) also are applicable to OSs (Table 1). However, epidemiology professional organizations have not yet done so, likely because of perceived idiosyncratic features and objectives of observational versus experimental research (22). OSs use a much wider range of designs, methods, and data sources than do CTs (3), arguably making it difficult to comprehensively and meaningfully summarize all possible scenarios tailored to a fixed number of informational elements recorded in a registry. Nonetheless, this bottleneck does not apply to all OSs because a fair number are indeed registered in resources such as ClinicalTrials.gov, and researchers can tailor their information to fit a typical CT registry format.

Last, a critical challenge is the completeness and quality of the information posted on the registry. A recent study of registered OSs showed large differences in the quantity and quality of information provided among and within registries (23). Still, the willingness of investigators to register their studies should be applauded because readers are informed about the existence of an OS and can contact the investigators for further information.

### OS PROTOCOLS AND INVESTIGATORS

Epidemiological research on human subjects must comply with accepted ethical standards that demand approval of the study protocol by institutional review boards (IRBs) (17, 24), particularly when investigators plan to obtain personal information directly from participants or will otherwise use potentially identifiable personal information about them. If investigators have submitted an OS protocol that is approved by an IRB, there is little additional burden to register that protocol on a publicly available registry. Amendments may be registered as new knowledge becomes available, creating a publicly available audit trail.

If IRBs approve studies lacking fully defined prespecified hypotheses, there is little reason for not making protocols available through subsequent registration. Although controversial, registration is common in large patient cohorts and biobanks established to collect information for future use (25). These studies are likely to have explicit protocols for recruiting participants, collecting data, and transparent sharing of OS data will document primacy of investigators who first registered the idea for a given data set. The act of registering a protocol and study hypotheses can serve as a claim to the conceptual underpinnings of a study.

### Table 1. Rationale for registration of CTs.

The Ottawa statement is provided in (21). All rationales listed are applicable to OSs.

<table>
<thead>
<tr>
<th>Ethical</th>
<th>Scientific</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Respect the investigator-participant covenant to contribute to biomedical knowledge by making trial methods and results public</td>
<td>• Increase the reliability and availability of evidence on which health care decisions are based</td>
</tr>
<tr>
<td>• Provide global open access to information</td>
<td>• Improve trial participation</td>
</tr>
<tr>
<td>• Reduce unnecessary duplication of invested research resources through awareness of existing trials</td>
<td>• Increase opportunities for collaboration</td>
</tr>
<tr>
<td>• Assure accountability with regard to global standards for ethical research</td>
<td>• Ensure transparency of trial design and methods</td>
</tr>
<tr>
<td>• Enable monitoring of adherence to ethical principles and process</td>
<td>• Provide open review of protocols to improve trial quality and refine methods</td>
</tr>
<tr>
<td></td>
<td>• Provide means for identification and prevention of biased underreporting or over-reporting of research</td>
</tr>
<tr>
<td></td>
<td>• Accelerate knowledge creation</td>
</tr>
</tbody>
</table>

---

lecting samples, and what measurements are to be made. The list of measurements may be updated and expanded over time, even when future analyses are unanticipated. As a future goal, all analyses performed on the data can be recorded in a computer log. This would show the path followed in exploratory analyses and facilitate replication by providing the exact analytical codes applied in additional data and studies. Some scientific fields are experimenting with live streaming, in which analyses are shared in real time. This technology allows the scientific community to observe, as it happens, the evolution of a research project and its analyses (26).

**STAKEHOLDER SUPPORT**

Influential stakeholders may provide incentives to encourage OS registration. Although a few high-profile journals encourage registration (2), the International Committee of Medical Journal Editors (ICMJE) has explicitly stated that OSs do not require registration (27). Unless ICMJE changes its position or leading medical journals independently start to require prospective registration of OSs, recommendations for registration by individuals will have limited influence because the major incentive of “license to publish” is missing.

The World Medical Association (WMA) lagged behind the ICJME, the World Health Organization (WHO), and various regulatory agencies (28, 29) in supporting the mandatory prospective registration of CTs, introducing this requirement in the 2008 version of the Declaration of Helsinki (17). In the 2013 revision (30), the WMA took the lead in requiring the registration of “every research study involving human subjects,” thus requesting OS registration up front. Such a decision could have an important impact because the behavior of physicians in complying with the ethical principles of such a declaration will influence the mindset of other professionals involved in observational research.

The critical role that WHO played in the registration of CTs is well known (31), including establishment of the international standards for CT registries (32) and the International CT Registry Platform (ICTRP) (33). If the ICTRP’s explicit mission truly is to “ensure that a complete view of research is accessible to all those involved in health-care decision making” (32), and if six of the eight scientific, ethical, and moral reasons why WHO supports registration of CTs (33) are equally applicable to observational research (Table 2), then WHO should support mandatory registration for OSs. Implementation of the ICJME request, in 2005, to register CTs (34) revealed the need to involve key agents of the research process to boost study registration. As the only stakeholder that approves all human research, IRBs could require that all OSs be registered as a condition of ethics approval.

We propose that public agencies and charitable organizations that fund medical research require prospective registration of OSs as a condition for funding, analogous to requiring IRB approval before study initiation. Major funding agencies such as the U.S. National Institutes of Health, which demand protocols before funding research, could garner broad acceptance for requiring public registration of OSs before launching a study. The same requirement could apply for competing grant renewals with extensions to collect new data and perform new analyses. The decision by the UK National Institute for Health Research Health Technology Assessment Programme to withhold a proportion of a research grant until a study report has been submitted for publication (35) underscores the power wielded by funders. In a similar manner, funders may consider withholding funding until the OS protocol is registered in sufficient detail.

---

### Table 2. Scientific, ethical, and moral reasons to support prospective registration of CTs [WHO, (33)].

<table>
<thead>
<tr>
<th>Information</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>There is a need to ensure that decisions about health care are informed by all of the available evidence.*</td>
</tr>
<tr>
<td>2.</td>
<td>It is difficult to make informed decisions if publication bias and selective reporting are present.*</td>
</tr>
<tr>
<td>3.</td>
<td>The Declaration of Helsinki states that “every trial must be registered in a publicly accessible database before recruitment of the first subject.”†</td>
</tr>
<tr>
<td>4.</td>
<td>Improving awareness of similar or identical trials will make it possible for researchers and funding agencies to avoid unnecessary duplication.*</td>
</tr>
<tr>
<td>5.</td>
<td>Describing clinical trials in progress can make it easier to identify gaps in clinical trials research.*</td>
</tr>
<tr>
<td>6.</td>
<td>Making researchers and potential participants aware of recruiting trials may facilitate recruitment.*</td>
</tr>
<tr>
<td>7.</td>
<td>Enabling researchers and health-care practitioners to identify trials in which they may have an interest could result in more effective collaboration among researchers. The type of collaboration may include prospective meta-analysis.*</td>
</tr>
<tr>
<td>8.</td>
<td>Registries checking data as part of the registration process may lead to improvements in the quality of clinical trials by making it possible to identify potential problems (such as problematic randomization methods) early in the research process.</td>
</tr>
</tbody>
</table>

*Reasons applicable to OSs. †This wording has been broadened in the 2013 revision of the Declaration of Helsinki (30): “Every research study involving human subjects must be registered in a publicly accessible database before recruitment of the first subject.”

### Table 3. Sharing OSs with the scientific community. Access could be extended to a lay audience; however, the nature of the information will be most useful to scientists. Modified from Khoury et al. (26).

<table>
<thead>
<tr>
<th>Information type</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data set registration</td>
<td>Should be feasible to achieve in large scale; each data set registers the variables that it has collected and their definitions; this would allow knowing how many studies with how many participants who have measured variables or markers of interest, instead of guessing what data are available on that marker beyond what has been published</td>
</tr>
<tr>
<td>Availability of detailed data</td>
<td>Individual-level (raw) data are made available; this practice may be subject to policy/consent/privacy constraints for past studies and their data; easier to anticipate and encourage in the design of future studies</td>
</tr>
<tr>
<td>Availability of data, protocols, and analyses codes</td>
<td>Optimal ability to evaluate the reproducibility of analyses, to maximize the integration of information across diverse studies, and to allow improvements on future studies based on exact knowledge of what was done in previous studies</td>
</tr>
<tr>
<td>Live streaming of analyses</td>
<td>Investigators not only post all their data and protocols online, but analyses are done and shown in real time to the wider community as they happen. Live streaming can be coupled with crowd sourcing of analyses across large communities of analysts</td>
</tr>
</tbody>
</table>
DATA SHARING

Beyond protocol registration for OSs, sponsors and investigators should consider sharing more OS information with the scientific community through a central repository in order to avoid fragmenting information among multiple registries (Table 3). Minimum information should include the study design, mode of data collection and outcomes, number of participants, and a full list of available measurements.

At the other end of the information-sharing spectrum, one could publicly disclose raw anonymized data at the individual-participant level and update the information as new data become available, provided informed consent stipulations have been pre-arranged and met. Ideally, raw data would be registered along with analysis codes. For some types of research, such as microarray analyses, most high-profile journals already require raw data, protocols, and analytical algorithms to be deposited in public platforms as a prerequisite to publication. This practice of data disclosure is becoming widespread (36). These examples suggest that the principle of having journals spearhead registration and public transparency is also feasible for OSs, as it has been for CTs (10). Eventually, both journals and funders may need to apply pressure to achieve desirable levels of transparency.

If, as Hernán and Wilcox (37) have stated, all study data “unless destroyed or lost, will one day be in the public domain,” why should the key information describing an OS protocol not be publicly disclosed by means of prospective registration? Funders are pivotal to this protocol not be publicly disclosed by means of prospective registration? Funders are pivotal to this as it has been for CTs (10). Eventually, both journals and funders may need to apply pressure to achieve desirable levels of transparency.

REFERENCES AND NOTES

4. T.L. Lash, Preregistration of study protocols is unlikely to improve the yield from our science, but other strategies might. Epidemiology 21, 612–613 (2010).
Making Prospective Registration of Observational Research a Reality
Rafael Dal-Ré, John P. Ioannidis, Michael B. Bracken, Patricia A. Buffler, An-Wen Chan, Eduardo L. Franco, Carlo La Vecchia and Elisabete Weiderpass

Sci Transl Med 6, 224cm1224cm1.
DOI: 10.1126/scitranslmed.3007513