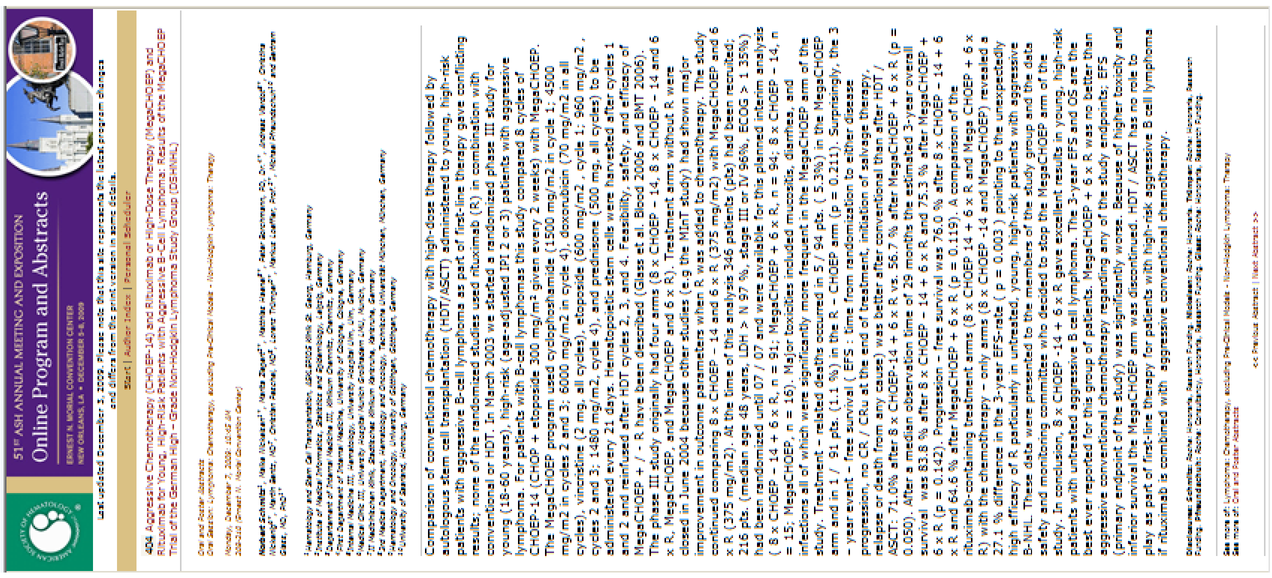
Let's Standardize Reporting of Clinical Trial Results! – A Patient Advocate's Perspective

An Opportunity to Improve Public Understanding and Trust in Clinical Research, while Making Analysis by Experts More Efficient

Reporting bias represents a major problem in the assessment of health care interventions.[1] Here we submit that the lack of standards on what needs to be reported and how it is formatted contributes to reporting bias, misunderstanding, and loss of trust in clinical research.

Pick a clinical abstract at random and you may get a feel for how challenging it can be to extract the key information and its significance:


51st ASH ANNUAL MEETING AND EXPOSITION
Online Program and Abstracts
ERNEST N. MORIAL CONVENTION CENTER
NEW ORLEANS, LA • DECEMBER 5-8, 2009

Last updated December 3, 2009. Please note that this site represents the latest program changes and differs from the print version in some details.

Start | Author Index | Personal Scheduler

404 Aggressive Chemotherapy (CHOEP-14) and Rituximab or High-Dose Therapy (MegaCHOEP) and Rituximab for Young, High-Risk Patients with Aggressive B-Cell Lymphoma: Results of the MegaCHOEP Trial of the German High – Grade Non-Hodgkin Lymphoma Study Group (DSHNHL)

Comparison of conventional chemotherapy with high-dose therapy followed by autologous stem cell transplantation (HDT/ASCT) administered to young, high-risk patients with aggressive B-cell lymphoma as part of first-line therapy gave conflicting results; none of the randomized studies used rituximab (R) in combination with conventional or HDT. In March 2003 we started a randomized phase III study for young (18-60 years), high-risk (age-adjusted IPI 2 or 3) patients with aggressive lymphoma. For patients with B-cell lymphomas this study compared 8 cycles of CHOEP-14 (CHOP + etoposide 300 mg/m² given every 2 weeks) with MegaCHOEP. The MegaCHOEP program used cyclophoshamide (1500 mg/m2 in cycle 1; 4500 mg/m2 in cycles 2 and 3; 6000 mg/m2 in cycle 4), doxorubicin (70 mg/m2 in all cycles), vincristine (2 mg, all cycles), etoposide (600 mg/m2 , cycle 1; 960 mg/m2 , cycles 2 and 3; 1480 mg/m2, cycle 4), and prednisone (500 mg, all cycles) to be administered every 21 days. Hematopoietic stem cells were harvested after cycles 1 and 2 and reinfused after HDT cycles 2, 3, and 4. Feasibility, safety, and efficacy of MegaCHOEP + / - R have been described (Glass et al. Blood 2006 and BMT 2006). The phase III study originally had four arms (8 x CHOEP – 14, 8 x CHOEP – 14 and 6 x R, MegaCHOEP, and MegaCHOEP and 6 x R). Treatment arms without R were closed in June 2004 because other studies (e.g the MInT study) had shown major improvement in outcome parameters when R was added to chemotherapy. The study continued comparing 8 x CHOEP – 14 and 6 x R (375 mg/m2) with MegaCHOEP and 6 x R (375 mg/m2). At the time of this analysis 346 patients (pts) had been recruited; 216 pts. (median age 48 years, LDH > N 97 %, stage III or IV 96%, ECOG > 1 35%) had been randomized until 07 / 07 and were available for this planned interim analysis (8 x CHOEP – 14 + 6 x R, n = 91; MegaCHOEP + 6 x R, n = 94; 8 x CHOEP – 14, n = 15; MegaCHOEP, n = 16). Major toxicities included mucositis, diarrhea, and infections all of which were significantly more frequent in the MegaCHOEP arm of the study. Treatment – related deaths occurred in 5 / 94 pts. (5.3%) in the MegaCHOEP arm and in 1 / 91 pts. (1.1 %) in the R – CHOEP arm (p = 0.211). Surprisingly, the 3 – year event – free survival (EFS : time from randomization to either disease progression, no CR / CRu at the end of treatment, initiation of salvage therapy, relapse or death from any cause) was better after conventional than after HDT / ASCT: 71.0% after 8 x CHOEP-14 + 6 x R vs. 56.7 % after MegaCHOEP + 6 x R (p = 0.050). After a median observation time of 29 months the estimated 3-year overall survival was 83.8 % after 8 x CHOEP – 14 + 6 x R and 75.3 % after MegaCHOEP + 6 x R (p = 0.142). Progression – free survival was 76.0 % after 8 x CHOEP – 14 + 6 x R and 64.6 % after MegaCHOEP + 6 x R (p = 0.119). A comparison of the rituximab-containing treatment arms (8 x CHOEP 14 + 6 x R and Mega CHOEP + 6 x R) with the chemotherapy – only arms (8 x CHOEP -14 and MegaCHOEP) revealed a 27.1 % difference in the 3-year EFS-rate (p = 0.003) pointing to the unexpectedly high efficacy of R particularly in untreated, young, high-risk patients with aggressive B-NHL. These data were presented to the members of the study group and the data safety and monitoring committee who decided to stop the MegaCHOEP arm of the study. In conclusion, 8 x CHOEP -14 + 6 x R gave excellent results in young, high-risk patients with untreated aggressive B cell lymphoma. The 3-year EFS and OS are the best ever reported for this group of patients. MegaCHOEP + 6 x R was no better than aggressive conventional chemotherapy regarding any of the study endpoints; EFS (primary endpoint of the study) was significantly worse. Because of higher toxicity and inferior survival the MegaCHOEP arm was discontinued. HDT / ASCT has no role to play as part of first-line therapy for patients with high-risk aggressive B cell lymphoma if rituximab is combined with aggressive conventional chemotherapy.

<< Previous Abstract | Next Abstract >>

Information is organized in a random order, and selected by the authors

Here we propose the use of required elements and a tabular format for reporting on trials for interventions against cancers and perhaps other life-threatening conditions – which we expect could be complementary to the National Institutes of Health goal to expand the Clinical Trials Registry and Results Database.[2]

Elements	TELL Report Format (proposed starting point):
N Evaluated / Intent to Treat	Number of participants in the clinical trials - Evaluated / Intent to Treat. We propose that intent to treat be included in all clinical research abstracts, expressed as: N = Evaluated / ITT Example: Evaluable: N = 300/500
Population Clinical circumstance	Medical condition (and subtypes): Risk: High, medium, low risk Performance index | Prognostic index Median number of prior therapies (and type) Median age | Genetic characteristics if any
Study Type	Phase: Randomized / Single arm Prospective / subset analysis
Primary Clinical Questions	Endpoints: Safety | Overall response rate | CR rate, Progression Free Survival, | Survival ... Provide pre-specified goal (relative to historical control) if a single-arm study
Primary Findings Met? Yes/No/Mixed	As defined in Primary Clinical Questions Expressed as Rate, include Confidence range, such as: Evaluated: CR/n (%) (CI range) Intent to Treat: CR/n (%) (CI range)
Secondary Clinical Questions	Endpoints: Safety | Overall response rate | CR rate, Progression Free Survival, | Survival ... Provide pre-specified goal (relative to historical control) if a single-arm study
Secondary Findings Met? Yes/No/Mixed	As defined in Secondary Clinical Questions Expressed as Rate, include Confidence range, such as: Evaluated: CR/n (%) (CI range) Intent to Treat: CR/n (%) (CI range)
Follow-up	Median follow-up: Final or Next:
Administration	How protocol was scheduled and administered Cycle = x, Number of Cycles, Number of Treatment Days, Treatment Duration in weeks Route: Oral, IV, Continuous infusion, Subcutaneous)
How Endpoints were measured	Summary of how outcomes were measured, such as: By: Independent / Investigator Schedule (weekly, monthly): Type (blood, imaging):
Maturation of Data	Completed / Interim? Time to enrollment and analysis? Median time of follow up: Need for follow-up?
Safety Results	Expressed as rate with range: By grade (severity): Serious first. For Evaluated: SE/n (%) (CI range) For ITT – if toxicities led to dropping out
Mortality	Death rate: Treatment-related: , Other: On study | Off Study Evaluated | Intent to Treat Expected rate in this population:
Limitations	Authors describe limitations of the study methods and design – such as sample size, or study type ... to describe level of evidence and if findings are consistent with other studies
Discussion	Free text area. Authors might provide here the implications of the findings – interpretations, and background that does not fit in the clinical results fields.

Information is organized logically as determined by peer review

Scientists may argue that the intended audience for clinical research is not the patient community or the public at large – that clinical abstracts are informal summaries (conversations) by and for scientists and that structured reporting would be burdensome. We remind that by definition clinical research requires patient participation – involves individuals who take substantial risks

---

[1] Natalie McGauran, et; **Reporting bias in medical research - a narrative review** ; Trials. 2010; 11: 37. 2010 April 13. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2867979/

[2] NIH: **Expansion of the Clinical Trials Registry and Results Data Bank** http://grants.nih.gov/grants/guide/notice-files/not-od-09-077.html

when participating in drug trials; and that improving patient outcomes is the core objective of clinical research.

We note also that there are many hidden costs, inefficiencies, and missed opportunities associated with free-form clinical reporting, including a substantial invitation for biased reporting – by what is left out, over-emphasized, or lost in the clutter.

> *"Ethical clinical research should contribute to generalizable knowledge and improve human health. The dedication of patients who take the risks to participate in clinical research is dishonored when their data remain secret."* - Alastair J.J. Wood, M.D [3]

Here we are compelled to add to Dr. Wood's comment that patients are short-changed in more subtle ways when the data is obscured or biased, even if unintentionally by how it is presented or interpreted by the various stakeholders. The **media** – tending to report in ways that attract readers, or the **drug sponsors** – tending to report in ways that attract investors, or the **investigators** – tending to report in ways that will support their hypothesis or enhances the significance of their work. This is not to suggest that evil intent or deliberate calculations to deceive are guiding the actions of investigators or drug sponsors, who truly play vital roles in clinical research.

Most patients when initially diagnosed with a cancer have little or no medical background or training in drug assessments or scientific method. Nor do we often have access to the full text of reports published in medical journals. However, the abstracts describing this research are widely available on the Internet or indirectly reported through press releases, which have a very poor track record for objective reporting.[4]

The review of clinical reports is a complex task that requires extensive training and skill. However, many patients facing life-threatening disease, or their loved ones, having an urgent need to know – will often do their best to uncover what the studies suggest or seem to prove in order to make more informed clinical decisions.

Faced with media-born misinformation and conflicting interpretations even among professionals, the public may lose trust in the clinical research process. Lacking standards for reporting, patient and physician analysis will be based often on incomplete information. The beliefs of patients will be based on happenstance – acquired from untrained parents, an influential friend, the claims made in shock media, a best-selling book or popular website – which makes the goal of informed medical decision-making more challenging than it needs to be. Amid the chaotic reporting standards we have observed that one report can be, unwisely, considered equivalent to any other. That is, patients may trust specific clinical trial reports too much or too little, or embrace them too selectively … based on what we want to be true, or based on the faith we have in certain individuals or institutions – or we may unwisely mistrust any study funded by a drug company or the government.

To help to address the confusion and its associated costs (bias, misinformation, and mistrust among them), we ask if a structured format could be required for clinical reports submitted to the medical journals – with a focus on the elements of clinical research that are key to assessment of any drug by the FDA for marketing approval. Noting that one need not have a deep understanding of the biology of the disease, or the mechanisms of a drug to appreciate which studies provide strong or weak evidence of meaningful clinical benefit if the key findings are reported consistently, and background information is provided about each of the key elements.

In **Table 1** which follows, we provide our draft proposal:
a **T**abular format with required outcome **E**lements in **L**ogic **L**ocations (TELL).

In **Table 2**, we propose reader-friendly but concise explanations of TELL elements for the public and the media. Each should be considered starting points or suggestions from an interested third party: patients!

To journal editors who may worry about the space requirements of a TELL-like format and the associated costs, we note that improving the clarity and objectivity of clinical reporting seems an excellent tradeoff. Further, free-form abstracts might still be used by journals if the abstract or full paper includes a link to a TELL. The full text of the published paper could then provide the technical illustrations and in-depth background for scientists, on the biology of the diseases and mechanisms of actions, which would continue to nurture productive conversations and support continuing progress against human disease.

Advantages of Standard Reporting in a TELL-like format:

- Enhance the ability of the scientific community to efficiently filter and weigh reports, and compare results across different studies and journals.

---

[3] Wood, Alastair J.J. **Progress and Deficiencies in the Registration of Clinical Trials**
N Engl J Med 2009 360: 824-830

[4] Daniel M. Cook et al; **Reporting Science and Conflicts of Interest in the Lay Press** PLoS ONE. 2007; 2(12): e1266.
http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2092382/

- Help sponsors and clinical investigators to make better decisions when designing clinical studies – to measure TELL events.

- Act as a deterrent against intentional or unintentional sponsor/investigator bias and common media-born misinformation.

- Discourage reporting of clinical data that has not yet matured.

- Help build support and improve public confidence in the objectivity of clinical science, needed to merit public funding of NIH.

- Provide a stronger basis for informed consent among patients and their treating physicians when considering clinical trials based on preliminary evidence of efficacy and safety.

- Foster more objective judgments among financial investors about which candidate drugs have the most potential, helping to attract needed capital to the more deserving inventions, while letting the less promising agents fail faster.

- By providing universal templates for abstracts the authors may produce higher quality abstracts more efficiently.

- And, as noted, such reporting would be complementary to the NIH initiative to expand the Clinical Trials Registry and Results Database. The results could be efficiently ported, from one registry to another if it is first reported in a structured way.

Finally, we have observed in FDA drug advisory committee reviews that the review is usually based on outcome events and detail about the study population (the context), as included in TELL below. Such information is rarely if ever proprietary – in need of protection from public disclosure.

Karl Schwartz
President, Patients Against Lymphoma
www.lymphomation.org

See TEL Tables 1 and 2 below.

Table 1

Elements	TELL Report Format (proposed starting point):
N **Evaluated / Intent to Treat**	Number of participants in the clinical trials - Evaluated / Intent to Treat. We propose that intent to treat be included in all clinical research abstracts, expressed as: N = Evaluated / ITT Example: Evaluable: N = 300/500
Population Clinical circumstance	Medical condition (and subtypes): Risk: High, medium, low risk Performance index | Prognostic index Median number of prior therapies (and type) Median age | Genetic characteristics if any
Study Type	Phase: Randomized / Single arm Prospective / subset analysis
Primary Clinical Questions	Endpoints: Safety | Overall response rate | CR rate, Progression Free Survival, | Survival … Provide pre-specified goal (relative to historical control) if a single-arm study
Primary Findings Met? Yes/No/Mixed	As defined in Primary Clinical Questions Expressed as Rate, include Confidence range, such as: Evaluated: CR/n (%) (CI range) Intent to Treat: CR/n (%) (CI range)
Secondary Clinical Questions	Endpoints: Safety | Overall response rate | CR rate, Progression Free Survival, | Survival … Provide pre-specified goal (relative to historical control) if a single-arm study
Secondary Findings Met? Yes/No/Mixed	As defined in Secondary Clinical Questions Expressed as Rate, include Confidence range, such as: Evaluated: CR/n (%) (CI range) Intent to Treat: CR/n (%) (CI range)
Follow-up	Median follow-up: Final or Next:
Administration	How protocol was scheduled and administered Cycle = x, Number of Cycles, Number of Treatment Days, Treatment Duration in weeks Route: Oral, IV, Continuous infusion, Subcutaneous)
How Endpoints were measured	Summary of how outcomes were measured, such as: By: Independent / Investigator Schedule (weekly, monthly): Type (blood, imaging):
Maturation of Data	Completed / Interim? Time to enrollment and analysis? Median time of follow up: Need for follow-up?
Safety Results	Expressed as rate with range: By grade (severity): Serious first. For Evaluated: SE/n (%) (CI range) For ITT – if toxicities led to dropping out
Mortality	Death rate: Treatment-related: , Other: On study | Off Study Evaluated | Intent to Treat Expected rate in this population:
Limitations	Authors describe limitations of the study methods and design – such as sample size, or study type … to describe level of evidence and if findings are consistent with other studies
Discussion	Free text area. Authors might provide here the implications of the findings – interpretations, and background that does not fit in the clinical results fields.

TABLE 2

Elements	These are proposed explanations of TELL elements, which to save space would not be included in published abstracts but could be available on the Internet for the public and media.
N Evaluated / Intent to Treat An easy way to judge the power of the study at a glance.	Stands for the number of participants in a study. It provides the denominator - a way to estimate the rate of results in the real world. To illustrate by extreme example, imagine how little confidence we can have in a study of two patients, reporting a 100% response rate. A meaningful denominator is absent from case reports and testimonials – a reason such reports are described as anecdotal – which is shorthand for not evidence of causality (that the intervention led to the result) or predictive of outcomes for others. Study results from a pre-defined N (or prospectively defined patient sample) provide more confidence than a numbered determined by chance, circumstances, or investigator ad hoc decisions. The latter could be determined when the outcome is most favorable, which undermines the integrity of the result.
Intent to Treat (ITT)	This number accounts for **all of the participants** that enrolled in the study, not just those who completed the protocol and were available for evaluation. When the ITT is greater than the number Evaluated, it calls into question the integrity of the analysis, and how well it could apply to results in the real world.
Population (Clinical circumstance)	How scientists and regulators interpret the results of a study is dependent on the population – the natural history of the disease untreated, or treated differently, but also the characteristics of the participants (age, performance, number and type of prior therapies). Did the study population have low or high-risk disease? For example, response rates in the previously untreated lymphoma patients can be more difficult to interpret than in those who have received many prior therapies.
Study Type	Randomized studies provide the most objective basis for identifying and comparing risks and benefits, relative to the control therapy – typically the standard of care.
Primary Clinical Questions	Endpoints describe what is being measured to determine if the intervention provided meaningful clinical benefit – net benefit or harm.
Endpoints	Of the measures used in clinical research, Survival is considered the most reliable as it accounts for measured and unmeasured effects. However, survival differences cannot always be measured for conditions that have a long clinical course, especially where other treatments will confound assessment … was it improved by the first or last treatment?
Methods: Protocol	Patients will want to know how the drug is administered: orally, by IV, by continuous infusion, and the duration of treatment.
Methods: Assessment	Notably, Independent data monitoring is often used in pivotal phase III trials to guard against biased interpretations, and to provide consistent evaluation methods.
Maturation of outcomes	Even after a study has completed the administration phase, many months or years may be needed to measure the endpoints, such as time to progression or other events being measured in the study. (A reason that validated biomarkers that predict longer-term outcomes are urgently needed to accelerate progress.)
Efficacy and Safety Results	To reliably calculate the response rates in the study population requires a pre-defined defined denominator (N), which is the basis for estimating the rates for study drug effects in the general population. Notably, case reports and testimonials, lacking a denominator (the number of participants), cannot be used to determine if the intervention even caused the outcome, or how likely the reported outcome will occur in others – which is critical to medical decision making.
Mortality Limitations	Mortality events can be acceptable in a population with high-risk disease. Reproducibility is the cornerstone of confidence in clinical outcomes – the objective assessment of risks and benefits. Size (N) counts, but having a second group achieve similar findings makes error (false negatives or positives) less likely. Randomized studies protect against patient selection bias and provide a reliable control to compare benefits and risks.
Discussion	Experts have noted that the conclusions of research authors are prone to bias, which can be considered a conclusive finding by the general public when published or quoted by the press.