The State of the Oncology Clinical Trials Portfolio: Insights from ClinicalTrials.gov

Methods

- A dataset comprising 91,346 clinical trials was downloaded from ClinicalTrials.gov on September 27, 2010 in XML format and a database for the Aggregate Analysis of ClinicalTrials.gov (AAGC) was created to facilitate analysis.
- A subset of trials was identified, corresponding to the FDA enactment of mandatory registration in 2007.
- A process was developed to annotate and validate disease conditions in order to create specialty datasets. A combination of National Library of Medicine (NLM) MeSH terms and non-MeSH (free-text) terms were used to do so as follows:
  - MeSH condition terms were identified using the 2010 MeSH thesaurus.
  - Non-MeSH condition terms (free-text) which appeared in five or more interventional studies registered after September 2007 were also identified.
- MeSH and non-MeSH terms were reviewed by clinicians and faculty within each clinical discipline at the Duke University Medical Center in order to annotate trials by disease.
- Next, the disease annotations provided by the clinicians and faculty were combined with MeSH condition terms generated by the NLM algorithm to generate a summary algorithm that categorized trials by disease specifically, as outlined in Figure 1.
- Trials identified as "oncology" were manually reviewed by clinicians to exclude false-positive studies and further classify the oncology trials by cancer subtype.

Results

- Of 40,970 interventional studies registered between October 2007 and September 2010, 8942 (22%) focused on oncology, the highest among all sub-specializations represented.
- In comparing oncology trials to those in other specialties, oncology trials were more likely to be single-arm (62% vs. 24%, p<0.001), open label (88% vs. 47%, p<0.001), and nonrandomized (64% vs. 23%, p=0.001).
- Oncology trials were also more likely to be early-phase (24% phase I or II in oncology vs. 5% in non-oncology).
- There was moderate but significant correlation between number of trials conducted by cancer subtype and associated U.S. incidence and mortality (incidence: correlation 0.56, P=0.037; mortality correlation 0.77; P=0.001).
- Only 63% of trials in oncology have a North American study-site. Among the top ten cancer types by incidence, less than half have the majority of their trials conducted only in North America.

Conclusions

- These data identify strengths and weaknesses in trial design, patient populations, and evidence development that need to be carefully considered in an era of increasing focus on research design and comparative effectiveness research.
- Subsequent analyses by CTTI will focus on subdividing these results by cancer type and impact of trial sponsor on the portfolio, to identify opportunities for improving the evidence development process in cancer.

Background

ClinicalTrials.gov is one of the largest databases of clinical research, comprising over 120,000 trials in 175 countries. With over 50 million page views a month, it is also the most utilized source for clinical trial information worldwide.

The database was initially created as a result of the US Food and Drug Administration (FDA) Modernization Act of 1997, the FDA made trial registration a requirement as of 2007 for all new clinical trials expected to contribute to a FDA submission. In addition, the International Committee of Medical Journal Editors mandates that all trials be included in a public registry as a requirement for publication of results in peer-reviewed medical journals.

Through a collaboration between the FDA and Duke University, as part of the Clinical Trials Transformation Initiative (CTTI), the goal of this project is to systematically summarize the relevant information in the ClinicalTrials.gov database to understand the full portfolio of clinical trials and the information in the ClinicalTrials.gov database to facilitate analysis.

The State of the Oncology Clinical Trials Portfolio: Insights from ClinicalTrials.gov

Background
ClinicalTrials.gov is one of the largest databases of clinical research, comprising over 120,000 trials in 175 countries. With over 50 million page views a month, it is also the most utilized source for clinical trial information worldwide.

The database was initially created as a result of the US Food and Drug Administration (FDA) Modernization Act of 1997, the FDA made trial registration a requirement as of 2007 for all new clinical trials expected to contribute to a FDA submission. In addition, the International Committee of Medical Journal Editors mandates that all trials be included in a public registry as a requirement for publication of results in peer-reviewed medical journals.

Through a collaboration between the FDA and Duke University, as part of the Clinical Trials Transformation Initiative (CTTI), the goal of this project is to systematically summarize the relevant information in the ClinicalTrials.gov database to understand the full portfolio of clinical trials and the information in the ClinicalTrials.gov database to facilitate analysis.

Methods
- A dataset comprising 91,346 clinical trials was downloaded from ClinicalTrials.gov on September 27, 2010 in XML format and a database for the Aggregate Analysis of ClinicalTrials.gov (AAGC) was created to facilitate analysis.
- A subset of trials was identified, corresponding to the FDA enactment of mandatory registration in 2007.
- A process was developed to annotate and validate disease conditions in order to create specialty datasets. A combination of National Library of Medicine (NLM) MeSH terms and non-MeSH (free-text) terms were used to do so as follows:
  - MeSH condition terms were identified using the 2010 MeSH thesaurus.
  - Non-MeSH condition terms (free-text) which appeared in five or more interventional studies registered after September 2007 were also identified.
- MeSH and non-MeSH terms were reviewed by clinicians and faculty within each clinical discipline at the Duke University Medical Center in order to annotate trials by disease.
- Next, the disease annotations provided by the clinicians and faculty were combined with MeSH condition terms generated by the NLM algorithm to generate a summary algorithm that categorized trials by disease specifically, as outlined in Figure 1.
- Trials identified as "oncology" were manually reviewed by clinicians to exclude false-positive studies and further classify the oncology trials by cancer subtype.

Results
- Of 40,970 interventional studies registered between October 2007 and September 2010, 8942 (22%) focused on oncology, the highest among all sub-specializations represented.
- In comparing oncology trials to those in other specialties, oncology trials were more likely to be single-arm (62% vs. 24%, p<0.001), open label (88% vs. 47%, p<0.001), and nonrandomized (64% vs. 23%, p=0.001).
- Oncology trials were also more likely to be early-phase (24% phase I or II in oncology vs. 5% in non-oncology).
- There was moderate but significant correlation between number of trials conducted by cancer subtype and associated U.S. incidence and mortality (incidence: correlation 0.56, P=0.037; mortality correlation 0.77; P=0.001).
- Only 63% of trials in oncology have a North American study-site. Among the top ten cancer types by incidence, less than half have the majority of their trials conducted only in North America.

Conclusions
- These data identify strengths and weaknesses in trial design, patient populations, and evidence development that need to be carefully considered in an era of increasing focus on research design and comparative effectiveness research.
- Subsequent analyses by CTTI will focus on subdividing these results by cancer type and impact of trial sponsor on the portfolio, to identify opportunities for improving the evidence development process in cancer.