Poster presented at DIA 2017

For more information:
bizint.com/DIA2017

Software for Business Intelligence
BizInt Smart Charts
www.bizint.com

What's the Best Download Format from NIH ClinicalTrials.gov? Comparing Data Availability in Different Export Options

Objective

The ClinicalTrials.gov registry website (NIH) offers several data formats for downloading search results. We examine differences in data field availability between formats. We further examine differences in the presentation of data fields when those differences affect the use of the data.

ClinicalTrials.gov - New for 2017

On June 1, 2017, the NLM announced a major change to the ClinicalTrials.gov platform[1]. We subsequently repeated the evaluations in this presentation using the ClinicalTrials.gov/beta platform.

Note that on the new platform, there are now two XML exports. The first (which we refer to as the Brief XML) is found under the "Select File Format" drop down list. This content is equivalent to the contents of the Excel (csv or tsv) export. The text for each study detail field is the same (aside from one difference noted in the Results section). The structure of the fields, however, is represented with XML markup which allows the user (with suitable tools) much more flexibility in the display of the content.

The second XML export (which we refer to as the Full XML) is the same XML export that was available from ClinicalTrials.gov prior to June 2017. This format is now accessed by the "Download" button in the section under the "For Advanced Users" heading.

Comments in the Brief XML downloaded from the beta website indicate that additional study fields may be added to the Brief XML in the future.

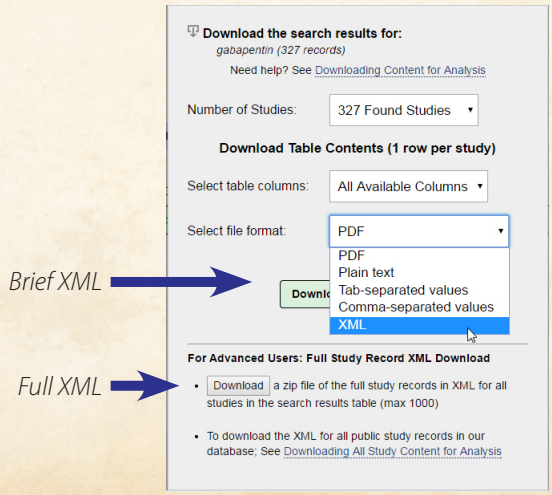
Method

We first compare field availability in each export format (CSV, Text, Brief XML, Full XML) for ClinicalTrials.gov based on field labels. Where labels are in a different format (XML element names), the closest match was identified.

The resulting matrix of available data elements is then validated by comparing actual clinical trial study details exported in each format. After confirming that the CSV, TSV, and Text exports held an identical list of fields with identical content, we selected the CSV format as the basis for all comparisons.

For each study field we attempted to map the data elements in each format required to achieve the display in the CSV export. Differences were noted.

The initial study was performed with an unconstrained search for 'gabapentin' (n=314). The study was reproduced on the new beta platform using an unconstrained search for 'nivolumab' (n = 436).



New ClinicalTrials.gov download panel

What's the Best Download Format from NIH ClinicalTrials.gov?

Results

As expected, the Full XML export of study details is the most comprehensive of the export formats. (The HTML display of study details on ClinicalTrials.gov is equally comprehensive, but only shows one study at a time). The only study detail field not present in the Full XML is the age groups, present as part of the Excel Age, and available as a separate element in the Brief XML format.

The text export formats (CSV, TSV, and TXT) all contain the same data elements. Furthermore, even though these file formats could allow different formatting within a data field, there are no differences between these formats within a field. The Brief XML has the same content as the text formats, but structured differently in each field.

The tables following show study detail fields found only in the Full XML (Table 1), those fields common to all formats (Table 2), and those fields with significant differences between formats (Table 3).

Table 1: *Study detail fields only found in Full XML.*

official_title
brief_summary (free text)
detailed_description (free text)
number_of_arms
arms_groups (label, type, description)
eligibility (free text with inclusion and exclusion criteria)
overall_official
overall_contact
locations (name, address, status, contact)
location_countries
reference (citations)
keywords
condition_browse (machine-assigned MeSH terms)
intervention_browse (machine-assigned MeSH terms)
clinical_results

Table 2: *Study detail fields present in all data formats.*

NCT Number
Other IDs
Acronym
Brief Title
Recruitment
Conditions
Gender
Phases
Sponsor/Collaborators [a]
Funded Bys [b]
Study Types
Study Designs
Enrollment [c]
First Received
Last Updated
Last Verified
Results First Received
Start Date
Completion Date [c]
Primary Completion Date [c]
URL

(a) Full XML differentiates lead sponsor, supporters.
(b) Full XML includes agency_class for each sponsor.
(c) Full XML indicates Actual vs Anticipated.

Table 3: *Study detail fields with significant differences between Full XML and the plain text formats.*

Age	Excel includes age range + age groups Brief XML has min/max age and groups Full XML has min/max age only
Interventions	Full XML includes additional descriptive details, arms group
Outcome Measures	Full XML differentiates between Primary and Secondary outcomes, includes time frame for each measure and descriptive details

Comparing Data Availability in Different Export Options

Results (Continued)

Additional Field Processing: Depending on the question being studied, additional processing may be required within Excel to extract data that is suitable for analysis, for example to separate data elements or remove labels. This additional processing (via formulae or macros) may add to the cost of using the CSV format.

The figure at right shows an extract from a TSV export, opened in Excel and formatted for alignment and text wrap.

By contrast, the figure at the bottom of the page shows a richer presentation of the same study details, based on a Full XML export of the same trials. This shows only a few variations possible with each field, due to limited space on this page.

Note that the Age field in Excel contains age groups, the one content item not found in the Full XML. These appear to be determined algorithmically (e.g. an age range of 18 and under is labeled as "Child, Adult")

B	G	J	L	Y
NCT Number	Interventions	Age	Enrollment	Outcome Measures
NCT02467361	Drug: BBI608|Drug: Ipilimumab|Drug: Nivolumab|Drug: Pembrolizumab	18 Years and older Â (Adult, Senior)	120	Determination of the safety and tolerability of BBI608 administered in combination with selected immunotherapeutic agent by assessing dose-limiting toxicities (DLTs)|Determination of the Recommended Phase 2 Dose (RP2D) by assessing dose-limiting toxicities (DLTs)|Assessment of the preliminary anti-tumor activity by performing tumor assessments every 8 weeks (Phase 2 portion)|Pharmacokinetic profile of BBI608 administered in combination with the selected immunotherapeutic agent as assessed by maximum plasma concentration and area under the curve|Pharmacodynamic activity of BBI608 administered in combination with the selected immunotherapeutic agent as assessed by biomarker analysis
NCT02073123	Drug: Indoximod|Drug: Ipilimumab|Drug: Nivolumab|Drug: Pembrolizumab	18 Years and older Â (Adult, Senior)	56	Overall Incidence of Adverse Events as a Measure of Safety and Tolerability|Phase 2 Dosing|Overall Response Rate|Number of Participants with Adverse Events as a Measure of Safety and Tolerability|Overall Survival|Mechanisms of activity/resistance to IDO/CTLA-4 inhibitor therapy|Progression Free Survival|Disease control rate

In the Interventions field, we can see important details such as dosage and the fact that drugs are being tested in combination. These details are not visible in the Excel export.

Finally for Outcomes, we are able to differentiate between Primary and Secondary outcomes, see a detailed description of the Outcome Measure (as seen in the Primary Outcome column), or show Measures together with the observed Time Frame (as in the Primary Outcome Timing column).

Trial Identifier	Primary Outcome	Enrollment	Min Age	Max Age	Interventions				Primary Outcome Timing	
					Name	Type	Description	Group	Measure	Time Frame
BBI608-201CIT NCT02467361	Determination of the safety and tolerability of BBI608 administered in combination with selected immunotherapeutic agent by assessing dose-limiting toxicities (DLTs) Determination of the Recommended Phase 2 Dose (RP2D) by assessing dose-limiting toxicities (DLTs)	120 (Anticipated)	18 Years	N/A	BBI608	Drug	Patients in this trial will receive BBI608 at assigned dose-levels according to the study arm the patient is enrolled into. BBI608 Dose Level 1: 240 mg twice daily, Dose Level 2: 480 mg twice daily. The assigned dose of BBI608 will be administered twice daily with approximately 12 hours between doses.	Combo with Ipilimumab; Combo with Nivolumab; Combo with Pembrolizumab	Determination of the safety and tolerability of BBI608 administered in combination with selected immunotherapeutic agent by assessing dose-limiting toxicities (DLTs)	6 weeks
									Determination of the Recommended Phase 2 Dose (RP2D) by assessing dose-limiting toxicities (DLTs)	6 weeks
					Ipilimumab	Drug	Ipilimumab 3 mg/kg is administered intravenously over 90 minutes every 21 days for a total of 4 doses.	Combo with Ipilimumab		
					Nivolumab	Drug	Nivolumab 3 mg/kg is administered as an intravenous infusion over 60 minutes every 14 days.	Combo with Nivolumab		
					Pembrolizumab	Drug	Pembrolizumab 2 mg/kg is administered as an intravenous infusion over 30 minutes once every 21 days.	Combo with Pembrolizumab		
NLG2103 NCT02073123	Overall Incidence of Adverse Events as a Measure of Safety and Tolerability - Phase 1 component: Evaluate the safety (adverse events - type, incidence, severity, duration, causality and treatment intervention) of the combination of indoximod and ipilimumab when given concomitantly. The safety and tolerability ipilimumab followed by Indoximod will be assessed by listing the overall incidence of AEs [CONT.]	56 (Anticipated)	18 Years	N/A	Indoximod	Drug	Initial dose of 600mg BID by mouth with escalation planned to 1200mg BID by mouth		Overall Incidence of Adverse Events as a Measure of Safety and Tolerability	17 months
					Ipilimumab	Drug	Ipilimumab administered intravenously at 3 mg/kg every three weeks for a total of four doses.	Indoximod + Ipilimumab	Phase 2 Dosing	22 months
									Overall Response Rate	22 months
					Nivolumab	Drug	Nivolumab administered intravenously at 3 mg/kg every three weeks	Indoximod + Nivolumab		
					Pembrolizumab	Drug	Pembrolizumab administered intravenously at 2 mg/kg every three weeks	Indoximod + Pembrolizumab		

What's the Best Download Format from NIH ClinicalTrials.gov? Comparing Data Availability in Different Export Options

Results (Continued)

It is worth noting that the XML formats for study details will always require additional software to translate the structured data into a format usable in reports, analysis, or presentations. This software will convert the text with XML markup into a more familiar format. A software tool may also allow you to select which fields to display, and how to format data for your particular purposes.

For the examples provided in this study, we used BizInt Smart Charts for Clinical Trials to convert the Full XML from ClinicalTrials.gov into an easier to read format. This tool was used in another poster at this meeting[2] to compare ClinicalTrials.gov results with studies retrieved from the EUDRA CT platform.

The comma separated and tab delimited formats can be directly imported into Excel, but in this case you are limited to a fairly simple view of the study details, discarding a fair amount of content for simplicity and ease of use.

Conclusions

When reporting on the results of a single study, the browser display on ClinicalTrials.gov is an easy to use and comprehensive display. When performing an analysis of many studies, the download formats allow the user to export a large number of results.

Most users of ClinicalTrials.gov use one of the text export formats like CSV when creating a report or visualization of a collection of studies. A primary reason is the easy access to office suite tools like Microsoft Office.

As we have shown, a significant fraction of the study details are not available in text exports, but are available in the Full XML. Many concepts, such as inclusion and exclusion criteria and locations, are only available in the Full XML. Other concepts such as outcome measures and ages are present in both Full XML and CSV, but more detail is available in the Full XML. Study results are only available for export in Full XML format.

Researchers working with study details should be aware of the additional study details available in the Full XML format, and consider adopting tools capable of translating the Full XML into a format suitable for analysis and presentation.

References

[1] Wolf K, Ide N, Koufopoulos J, Williams RJ, Tse T. *ClinicalTrials.gov: First in a Series of Changes to Improve Usability for Stakeholders.* NLM Tech Bull. 2017 May-Jun;(416):e4

[2] Eberle M. Searching the registries for trial submission QA and competitor intelligence - Comparing ClinicalTrials.gov and EudraCT. DIA 2017 Professional Posters. 2017 Jun.

For more information and a free trial:

www.bizint.com/dia2017