



European Society
for Blood and Marrow
Transplantation

Live Survey - Interactive session on data collection issues

Beate Lindner

Lisbon, 19/03/2018

**No conflict of
interest**



***Data! Data! Data! I can't make
bricks without clay!***

*Sherlock Holmes (Sir Arthur Conan
Doyle)*



Data Managers and Physicians
Do they speak the same
language?

Data manager Dana to physician Doctor X.:

"Can you please help me with some questions. I would like to complete the documentation of my MED-A/B-records for patient Smith, and I cannot find all answers in the letters."

Doctor X: *"Please ask later, I am very busy at the moment."*

A few days later.

Same request

"Why do you not contact our new assistant, Dr Y?"

Dana tells her questions to the new assistant and realises from his answers that he cannot help her.

Later in the year, after several more attempts to get help – short time before a congress:

Doctor X.: *"Dana, could you please run some queries on our data ... I would need the following"*

She runs the queries and sends them to the physician:

... 'Dana, why do we have so many missing items?'

Live Survey _ Data collection

One of the aims of this session is to talk together – live survey! -
about our experiences when collecting data and sharing possible solutions and ideas for our everyday life.

Live Survey _ Data collection

In addition, I also would like to show a broader view on the whole business of data collection that hopefully is inspiring and motivating for us.

Live Survey _ Data collection

- What are the practical things and thoughts you would like to share with others?
- Welcome are small and big things that can make a change to our everyday data collection duties.

Live Survey _ Data collection

References and quotes in this presentation and a lot of the overall structure are owed to **Brent Dykes**

(author of Web Analytics Action Hero. He has been working in web analytics with enterprise clients for more than seven years at Adobe/Omniture.)

Quotes via the following two websites:

<http://www.analyticshero.com/2012/09/19/the-two-guiding-principles-for-data-quality-in-digital-analytics/>

and

<http://www.analyticshero.com/2012/10/25/31-essential-quotes-on-analytics-and-data/>

His ideas have been very inspiring for me, though coming from a completely different field.

“Regardless of whether you create, share, or consume data, you should be concerned about the quality of your [...] data within your organization.”

*“the two key principles for data quality are **usefulness** and **trust**”*

Usefulness

mainly depends on what our national authorities, our center, the EBMT decide on what needs to be collected (MED-A)

nevertheless, we
can give input

Usefulness

reduced by
missing items, wrong items, data entered
too late

Trust

- we must decide which data sources can be used and what can be reported
- we must decide when to seek for advice and assistance

Trust

- the data entry program can help by making data checks
- we can check (older) data by using data quality queries

Trust

- to keep in mind, too: we are responsible in our field in regard to the data privacy laws
(new regulations from May 2018)

Live Survey _ Data collection

Looking at

- Accuracy
- Consistency
- Relevance
- Completeness
- Timeliness

from a data manager's perspective or the person who needs to collect the data.

Accuracy

- Do we know what data we collect?
- How reliable are our data sources?
- Do we have the right definitions?
- Do the EBMT / our transplant center provide us with the necessary information?
- Do we get the training we need for collecting the data?

***„Greater accuracy
requires more
time and effort.“***

Consistency

- Which data guidelines and reporting principles do we have?
- Whom should we contact if we have questions ... regarding definitions (e.g. disease classifications, disease status, stage)
- Who informs us about changes in definitions?
- Where do we find them?

Consistency

- Is our data in line with overall findings? Do we have 'strange' data (e.g. unusual conditioning regimes, drug doses, procedures, definitions)
- Inbuilt tests in Promise/ more in MACRO should help

Relevance

- Do we find in our data sources the relevant data for the MED-A/-B data reports?
- If not, what can we do about it?
- Is it difficult and time consuming to collect these data?
- Can we change anything about this?

Completeness

- Can we answer all questions in the MED-A (-B) forms?
- Are there several cases – e.g. certain disease groups, patient groups, or treatments – where documentation is not complete?
- How can we ask for completeness in such cases?
Who can be of help for us?

Completeness

- Do we get too many data requests?
- When and where do we stop striving after completeness?

Timeliness

- When do we report the data?
- Can we manage reporting around day 0, day 100 and annually?
- What about old missing data queries?
- When do we stop looking for certain data?

Data quality

- iterative process
- means performing analysis on our data regularly in order to identify data quality problems
- Would we like to get more feedback from registries?

An **early warning system** ideally should be implemented in the database system.

It should help data managers keeping track of missing items, things that need to be checked, and (day 100/FU) reports due.

Some more ideas from the business world:

Feedback loops could be of enormous help to improve the whole field of data collection and data quality. This means team-work between data managers, physicians, data analysts and other persons involved in the data collection process.

Live Survey _ Data collection

*If you do not know how to ask the right question,
you discover nothing.*

W. Edward Deming

Live Survey _ Data collection

“While data is important, the right data is essential. It’s becoming easier to feel overwhelmed by the increasing amounts of data being collected. Understanding what’s important [...] helps analysts to evaluate what data counts or should be counted. Albert Einstein was on to something when he said we can’t always track what we want to count, but that doesn’t stop us from constantly exploring new ways to get the data we need (just remember to respect privacy policies and laws).”

Live Survey _ Data collection

“Maintaining the integrity of your [...] data is like managing a garden; it’s something that requires constant attention. If your implementation becomes overrun by weeds, you may lose your entire crop of useful data. All the hard work that went into planning and planting the crop will be wasted, and there won’t be a bountiful harvest of meaningful insights and action.”

Live Survey _ Data collection

Finally, we should be aware of our role and stress that we as data managers cannot manage everything (alone) and without the necessary support.

Facts do not cease to exist because they are ignored.
Aldous Huxley

Not everything that can be counted counts, and not everything that counts can be counted.

Albert Einstein