



Man vs. machine

Reflections on machine-assisted and human-driven approaches for examining open-text progress reports

Dr. Stephanie Quail



The challenge – understanding participant trajectories

- **The project:** ARTD's evaluation of the Queensland Drug and Alcohol Court
- **The question:** What distinguishes people who go on to have success in the program from those who start but don't graduate?
- **The limitations:**
 - Interviews with participants –rich source of info about journey, but difficult recruitment, only able to engage those with positive experiences
 - Administrative program data – includes ultimate outcomes and participant characteristics, but doesn't capture all key factors or within-program trajectory

A potential solution? Analysing open-text progress reports

The problem – wrangling open-text progress reports

- **Why should we analyse progress reports?**

- Progress reports can contain rich information about participants' experiences, and frequently describe theoretically important risk and protective factors that are not typically recorded in administrative datasets.

- **Why don't we typically analyse progress reports?**

- The unstructured narrative nature of progress reports, and the sheer volume of it, typically means that analysis of this type of open-text data is very time and resource intensive.

Is there a way to make the most of this type of data, within the time and resource constraints of an evaluation?

What we did – man, machine, in-between

- **What we wanted to know:**

- Can we identify different patterns in participant trajectory?
- What key factors are most important for participant success?

- **Our approach:** We took three different, but complementary, approaches :

1. Machine-based: Text-mining and sentiment analysis
2. Human-driven: Hand-coding key factors and sentiment
3. Hybrid: Qualitative Comparative analysis (QCA)

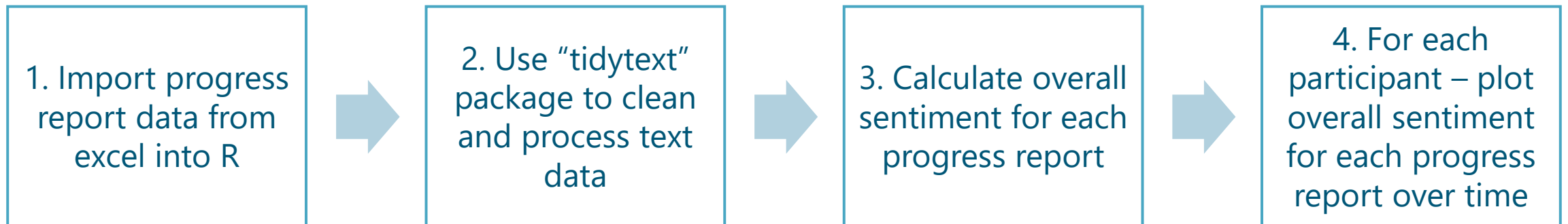
What were the benefits and challenges of these methods?
How do humans stack up against machine-assisted approaches?

1. Text-mining & sentiment analysis

- **What is it?**

- Text-mining: extracting useful information and patterns from large volumes of unstructured text
- Sentiment analysis: identifying and classifying the emotional tone or content of text

- **How did we do it?**



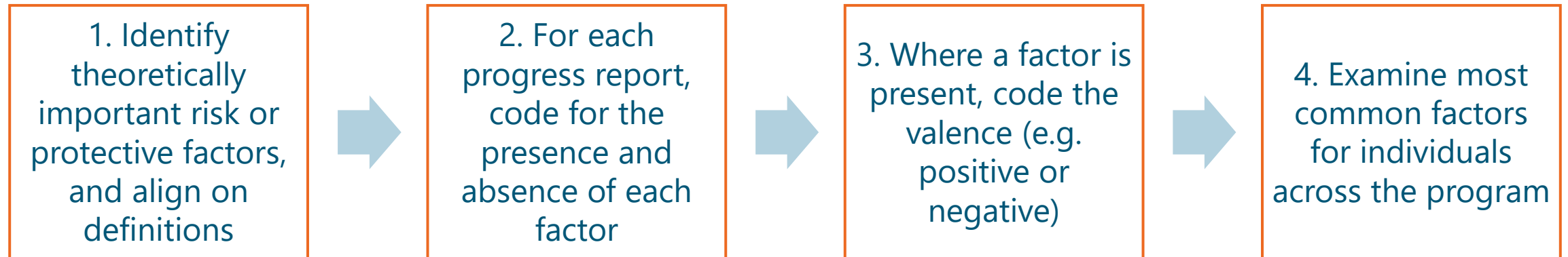
<https://www.tidytextmining.com/>

2. Hand-coding of key factors

- **What is it?**

- Team of humans using their understanding of the text and coding framework
- Coding for presence/ absence of factors, as well as valence of description of factor (e.g. positive - reference to supportive relationships vs. negative - reference relationship breakdown)

- **How did we do it?**

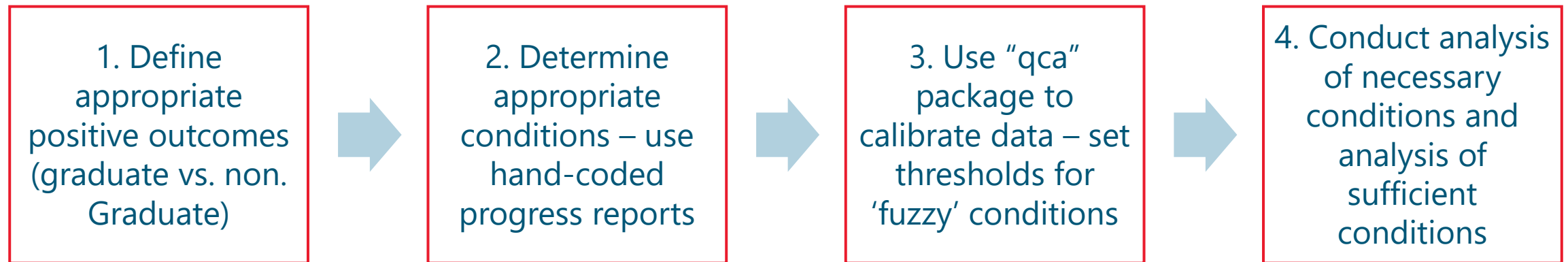


3. Qualitative comparative analysis (QCA)

- **What is it?**

- QCA blends qualitative and quantitative data to analyse causal patterns using a small to medium number of cases.
- Looks to identify *necessary* and *sufficient* configurations of conditions

- **How did we do it?**



<https://bookdown.org/dusadrian/QCAbook/>

What did we learn?

1	2	3
Text-mining and sentiment analysis	Hand-coding of key factors	QCA
<ul style="list-style-type: none">• Best for a quick, rough overview of what is in the data• Needs good knowledge of context to sense-check, difficult with more complex text	<ul style="list-style-type: none">• Best for nuance, complexity, or where there is uncertainty• With volume incredibly time intensive, inter-coder variability	<ul style="list-style-type: none">• Best of both worlds with small to medium sample sizes• Can be difficult to identify and define appropriate conditions, conduct, explain



Each approach has unique strengths and limitations. Matching what you need and what is most important with the appropriate technique will help you make the most out of open-text data

What to think about for next time?

1. What data is available?

- What is in the open-text data? Who has recorded it?
- Is it likely to be used in a consistent manner across staff/ sites?
- Can it be appropriately de-identified?

2. What are the time/ resource constraints?

- How much data is available? Do you need to look at all or just a sample?
- How quickly does this need to happen?
- What is the

3. How will the results of this analysis be used?

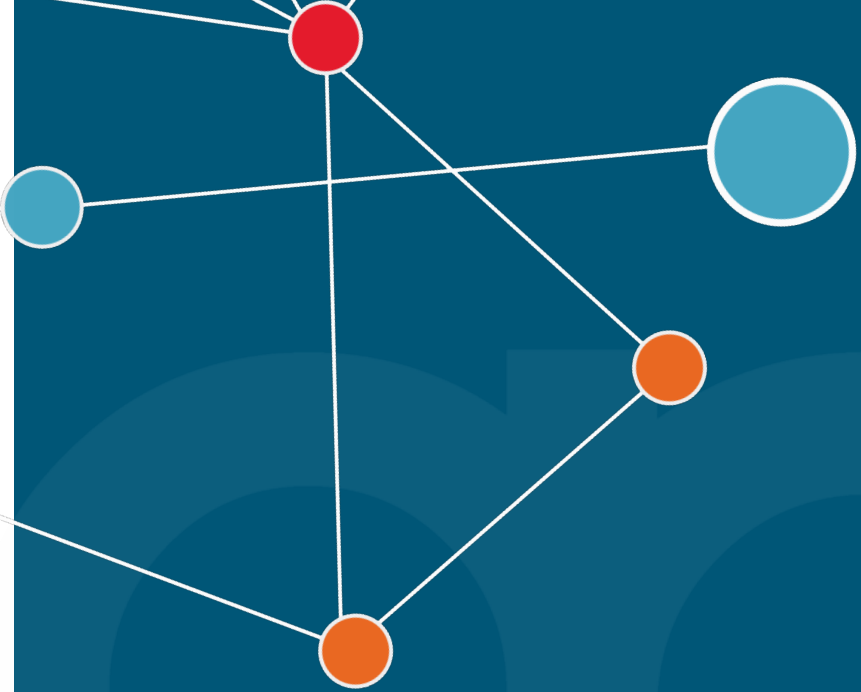
- Exploratory vs. confirmatory? Level of accuracy required?

Useful resources

- R for Data Science
 - <https://r4ds.had.co.nz/index.html>
- Text-mining with R
 - <https://www.tidytextmining.com/>
- QCA with R
 - <https://bookdown.org/dusadrian/QCAbook/>

Contact:
Stephanie Quail
Manager, ARTD Consultants
stephanie.quail@artd.com.au

Connect with us



Meanjin
Brisbane

Eora
Sydney

Narm
Melbourne

