MAPPING IMPLEMENTATION RESEARCH TRENDS IN THE WHO REGIONS TEXT MINING OF MOOC RESEARCH LETTERS OF INTENT

Progress report 28 March 2020

The objective of this project is to develop a method and analytical tools to automatically extract information from batches of research proposals submitted as part the final exam of the TDR MOOC on Implementation Research (IR). The findings consist of mapping the implementation research needs through interactive information products for public access (databases, maps, and dashboards) to promote IR as an essential prerequisite to effective public health interventions.

The text mining was performed on 581 letters of intent (LoIs) with evaluation scores of at least 70%. We extracted countries, illnesses, research methods and implementation research strategies and outcomes. The textual analysis was based on words and expressions to detect frequencies and co-occurrences of specific terms. The analysis was done in the programming language R with packages for text mining 'tm', 'quanteda' and for graphics 'ggplot2'. The results are displayed in a Tableau dashboard.

The results (frequencies) are presented in a Tableau dashboard that can be accessed online and the R script that outlines the method can be accessed through GitHub. A version of the dashboard that can be embedded in the TDR website with live feeds from an online form through which the LoIs are submitted is still in progress.

1.1 Corpus preparation
1. Import LoIs in pdf into R as a list
2. Remove text from references
3. Create corpus from text
4. Add document-level variables to the corpus

1.2 Textual processing
1. Compile lists of neglected tropical diseases (A) and countries (B) based on WHO and World Bank lists.
2. Create regular expressions for diseases and countries.
3. Extract the words and expressions based on detection techniques (including flowcharts and image captions).
4. Classify countries into world regions (C).
5. Classify diseases into wide topic areas (infectious diseases of poverty, sexual and reproductive health, non-communicable diseases).
6. Select countries and illnesses with frequencies above the average number of mentions per LoIs.
7. Manual input of missing cases for disease and country.
8. Manual validation of a random sample of LoIs.

1.3 Extract research methods
1. Develop a glossary of qualitative and quantitative methods based on the TDR's toolkit and the MOOC transcript (see Table 1).
2. Extract research methods from the LoIs using detection techniques.
3. Select methods with frequencies above the average number of mentions per LoIs.
4. Manual input of missing cases for research methods.
5. Manual validation of a random sample of LoIs.

1.4 Extract implementation research strategies
1. Identify a glossary of 73 IR strategies from the literature Powell et al (2015) – see Table 2.
2. Manual scan the transcript of the MOOC to select IR strategies.
3. Match the 73 research IR strategies identified in the literature with those identified in the MOOC transcript. This was done by two independent researchers that agree on 76% of the classification of IR strategies. The remaining 24% of strategies were discussed with a third researcher to reach a consensus.
4. Extract implementation research strategies based on expressions identified in the previous step
5. Extract IR outcomes based on synonymous and regular expressions (acceptability, adoption, appropriateness, cost, feasibility, fidelity, penetration, sustainability (see Table 3).
6. Manual input of missing cases for implementation research strategies and outcomes.
7. Manual validation of a random sample of LoIs.

Table 1: Glossary of Research Methods

General terms	Research Designs	Data sources	Qualitative methods	Quantitative Methods
• Methodology • Design • Research Design • Research Strategy • Research Approach • Methods • Literature Review • Policy Review • Situation analysis • Cost analysis/assessment	• Descriptive • Observational study • Exploratory design • Cross-sectional • Comparative • Analytical • Correlational • Intervention/Experimental • Adaptive design trial • Quasi-Experimental • Natural Experiment • Randomized Control Trial/RCT • Pragmatic trial • Pre-post design • Before-after research design • Interrupted time-series study • Cohort study • Longitudinal study • Case-control study • Mixed-Methods • Sample, sampling: randomised, randomized cluster, purposive, snowballing, step-wedged cluster,	• Routine program/intervention/service data/records • Treatment logs • Treatment records • Clinic records • Administrative Data • Data on service utilization • Program communication/outreach messages and data • Patient outreach messages and data • Social data • Primary data • Secondary data • Video/film/photographs • Program reports • Evaluation reports • Meetings • Diaries • Community meetings • Informal group discussion • Clinical processes • Stakeholder interactions	• Qualitative interviews • Individual interviews • Semi-structured interviews • In-depth interviews • Expert interviews • Key informant interviews KII • Narrative interviews • Participant observations/assessments • Ethnographic observations/assessments • Direct observation/assessment • Field observations • Transect walk • Focus group discussion/FGD • Document analysis • Public discourse • Discourse Analysis • Media discourse analysis	• Survey • Survey questionnaire • Structured survey • Face-to-face survey • Phone survey • Online self-administered survey/questionnaire • Self-administered survey/questionnaire • Household survey • Disease burden survey • Patient survey • Indicator survey • Demographic survey • Health survey • Information survey • Tally sheet • Content analysis • Skills assessment scorecards

	convenience, simple random, systematic random, stratified, successive, respondent driven, maximum variation, contrasting cases • Control group/Treatment group		• Policy discourse analysis • Document review • Data review	• Knowledge assessment scorecards • Checklist • Community scorecards • Document review • Data review

Table 2: Glossary for Implementation Strategies

- Access new funding
- Assess for readiness and identify barriers and facilitators
- Alter incentives/allowance structures
- Assess the readiness and identify barriers and facilitators
- Audit and provide feedback
- Centralized technical assistance
- Change record systems
- Conduct local needs assessment
- Conduct local needs assessment
- Conduct ongoing training
- Conduct ongoing training
- Create new clinical teams
- Develop and implement tools for quality monitoring
- Develop and organise quality monitoring systems
- Develop and organize quality monitoring system
- Develop effective materials
- Engage or include patients/consumers and families in the implementation effort
- Facilitate relay of clinical data to providers
- Facilitation
- Fund and contract for the clinical innovation
- Identify and prepare champions
- Involve executive boards
- Involve patients/consumers and family members
- Mandate change
- Promote network weaving
- Provide clinical supervision
- Provide clinical supervision
- Provide local technical assistance
- Purposely re-examine the implementation
- Develop and organize quality monitoring systems
- Stage implementation scale-up

• Stage implementation scale-up	
• Tailor strategies	
• Use mass media	
• Use other payment schemes	

Table 3: Implementation Research Outcomes

IR outcome	Other terms
Acceptability	Satisfaction
Adoption	Uptake, utilization
Appropriateness	Relevance, compatibility, suitability, usefulness, practicability
Cost	Marginal cost, cost-effectiveness, cost-benefit
Feasibility	Actual fit or utility, suitability for everyday use
Fidelity	Delivered as intended, adherence, integrity, quality of program delivery
Penetration	Level of institutionalization, spread, service access
Sustainability	Maintenance, continuation, incorporation, integration, routinization