FAIR scientific information with the Open Research Knowledge Graph

Markus Stocker
November 14, 2022
EOSC Symposium
More than in LV tissue samples from unused donor hearts (Figure 5A). As shown by immunofluorescence staining, TRPV1 expression was significantly reduced in failing hearts (most pronounced in patients with ischemic cardiomyopathy) (Figure 5B). Protein expression levels of the transient receptor were significantly lower in failing hearts than in the controls (Figure 5C).
failure than in LV tissue samples from unused donor hearts (Figure 1A). As shown by electrophoretic mobility shift assays, IRE binding activity was significantly reduced in failing hearts (most pronounced in patients with ischemic cardiomyopathy) (Figure 1B). Protein expression levels of the transferrin receptor were significantly lower in failing hearts than in the controls (Figure 1C).
failure than in LV tissue samples from unused donor hearts (Figure 1A). As shown by electrophoretic mobility shift assays, IRE binding activity was significantly reduced in failing hearts (most pronounced in patients with ischemic cardiomyopathy) (Figure 1B). Protein expression levels of the transferrin receptor were significantly lower in failing hearts than in the controls (Figure 1C).

Figure 1. Reduced IRE activity and iron content in failing human hearts. (A) Non-heme iron concentration in left ventricle (LV) tissue samples from non-failing donors (NF) and patients with cardiac failure (P<0.001 vs. non-failing donor; COX vs. dilated cardiomyopathy (DCM), n=4 per group). (B) Representative electrophoretic mobility shift assay and summary data showing non-irrelevant element (IRE) binding activity in LV tissue samples. (C) Representative immunoblot and summary data showing transferrin receptor 1 (TFR1) and GAPDH protein expression in LV tissue samples. n=4-6 (controls; no sample loaded).
failure than in LV tissue samples from unused donor hearts (Figure 1A). As shown by electrophoretic mobility shift assays, IRE binding activity was significantly reduced in failing hearts (most pronounced in patients with ischemic cardiomyopathy) (Figure 1B). Protein expression levels of the transferrin receptor were significantly lower in failing hearts than in the controls (Figure 1C).

![Image of a figure showing IRE binding activity](https://doi.org/10.1093/eurheartj/ehw333)
Student’s t-test [http://purl.obolibrary.org/obo/OBI_0000739]

has dependent variable
iron-responsive element binding [http://amigo.geneontology.org/amigo/term/GO:0030350]

has specified input
https://doi.org/10.4563/zenodo.56980

has specified output
p-value [http://purl.obolibrary.org/obo/OBI_0000175]
scalar value specification “0.000000013112475”^^xsd:decimal
“Scientific writing can [...] be called information burying”

First we bury it and then we mine it again

-- Barend Mons (2005)

https://doi.org/10.1186/1471-2105-6-142
“we have failed to [...] organize [...] information [...] in rigorous [...] ways, so that finding what we want and understanding what's already known become [...] increasingly costly experiences”

-- Teresa K. Attwood et al. (2009)

https://doi.org/10.1042/BJ20091474
“Despite recent developments in machine learning [...], data extraction is still largely a manual process”

-- Julian Higgins et al. (2022)

https://training.cochrane.org/handbook/current
The benefits are obvious
<table>
<thead>
<tr>
<th>Location</th>
<th>Time period</th>
<th>Basic reproduction number</th>
<th>Has value</th>
<th>Confidence interval (95%)</th>
<th>Lower confidence limit</th>
<th>Upper confidence limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lombardy, Italy</td>
<td>2020-01-14 - 2020-03-08</td>
<td>3.1</td>
<td>2.9</td>
<td>3.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Iran</td>
<td>2020-02-19 - 2020-02-29</td>
<td>3.6</td>
<td>3.4</td>
<td>4.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Iran</td>
<td>2020-02-29 - 2020-02-29</td>
<td>3.58</td>
<td>1.29</td>
<td>8.46</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Singapore</td>
<td>2020-01-21 - 2020-02-26</td>
<td>1.27</td>
<td>1.19</td>
<td>1.36</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```python
import requests
import datetime
import pandas as pd
import numpy as np
from orkg import ORKG
from bokeh.io import output_notebook
from bokeh.models import ColumnDataSource, HoverTool, WheelZoomTool, ResetTool, SaveTool, PanTool, DatetimeTickFormatter, Whisker
from bokeh.plotting import figure, show, output_notebook

output_notebook()

orkg = ORKG(host='https://orkg.org/orkg', simcomp_host='https://orkg.org/orkg/simcomp')

df = orkg.contributions.compare_dataframe(comparison_id='R44930')

dates = np.array([datetime.datetime.fromisoformat(x) for x in df.loc['has end', :]])
values = np.float32(df.loc['Has value', :])
lower = np.array([np.float32(x) if x else np.nan for x in df.loc['Lower confidence limit', :]])
upper = np.array([np.float32(x) if x else np.nan for x in df.loc['Upper confidence limit', :]])

hover1 = HoverTool(
    tooltips=[
        ('Date', '@date{%F}'),
        ('RQ', '@value{0.0f}'),
        ('95% CI', '@lower{0.0f}--@upper{0.0f}')
    ],
    formatters={
        '@date': 'datetime',
        '@value': 'printf',
        '@lower': 'printf',
        '@upper': 'printf'
    }
)

df = pd.DataFrame(data=dict(date=dates, value=values, lower=lower, upper=upper))
source = ColumnDataSource(df)

def figure(x_axis_type='datetime', y_range=(0, 9), plot_width=800, plot_height=350, tools=[hover1, WheelZoomTool(), PanTool(), ResetTool()], pxaxis.formatter=DatetimeTickFormatter(days=['%d %b']))
```

1 Work

COVID-19 Reproductive Number Estimates
Allard Oelen, Jennifer D'Souza, Markus Stocker, Lars Vogt, Kheir Eddine Farfar, Muhammad Haris, Kamel Fadel, Mohamad Yaser Jaradeh & Vitalis Wiens

Comparison published 2020 in Open Research Knowledge Graph (ORKG)

Comparison of published reproductive number estimates for the COVID-19 infectious disease

DOI registered October 16, 2020 via DataCite.

https://doi.org/10.48366/r44930
Transmission potential of COVID-19 in Iran
Posted Content published 2020 via medRxiv
We computed reproduction number of COVID-19 epidemic in Iran using two different methods. We estimated $R_0$ at 3.6 (95% CI, 3.2, 4.2) (generalized growth model) and at 3.58 (95% CI, 1.29, 8.46) (estimated epidemic doubling time of 1.20 (95% CI, 1.05, 1.44) days) respectively. Immediate social distancing measures are recommended.

Other Identifiers
Publisher ID: medRxiv:2020.03.08.20030643v1
DOI registered April 10, 2020 via Crossref.

2 Citations
https://doi.org/10.1101/2020.03.08.20030643

2 Citations
COVID-19 Reproductive Number Estimates
Allard Oelen, Jennifer D’Souza, Markus Stocker, Lars Vogt, Kheir Eddine Farfar, Muhammad Haris, Kamel Fadel, Mohamad Yaser Jaradeh & Vitalis Wiens
Comparison published 2020 in Open Research Knowledge Graph (ORKG)
Comparison of published reproductive number estimates for the COVID-19 infectious disease
DOI registered October 16, 2020 via DataCite.
So why is most scientific information still buried in documents?
Because it is hard to produce FAIR scientific information
<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic reproduction number</td>
<td>3.1</td>
</tr>
<tr>
<td>Location</td>
<td>Lombardy, Italy</td>
</tr>
<tr>
<td>Time period</td>
<td>2020-01-14 - 2020-03-08</td>
</tr>
<tr>
<td>Research problem</td>
<td>Determination of the COVID-19</td>
</tr>
</tbody>
</table>

Determination of the COVID-19 basic reproduction number

Referred: 35 times

Instance of: Problem
The suggestions listed below are automatically generated based on the title and abstract from the paper. Using these suggestions is optional.
# Two Linear Mixed Model (LMM) computations

```r
lm.mwd.1 <- lmer(MWD_cor ~ cc_variant + (1|depth), data = df.MWD)
lm.mwd.2 <- lmer(MWD_cor ~ cc_type + (1|depth), data = df.MWD)
```

# Output data for the two LMM
```r
df1 <- data.frame(summary(lm.mwd.1)$coefficients, check.names=FALSE)
df2 <- data.frame(summary(lm.mwd.2)$coefficients, check.names=FALSE)
```

```r
instance <- tpsmodel_fitting{
  label="Linear mixed model fitting with MWD as response, CC variant as predictor variable, and soil depth as random variable",
  has_input_dataset=tuple(df.MWD, 'Difference of mean weight diameter between the dry and wet sieving method'),
  has_input_model=tpsstatistical_model(
    label="A linear mixed model with MWD as response and CC variant as predictor variable",
    is_denoted_by=tpsformula(
      label="The formula of the linear mixed model with MWD as response and CC variant as predictor variable",
      has_value_specification=tpsvalue_specification(
        label="MWD_cor ~ cc_variant + (1|depth)",
        has_specified_value="MWD_cor ~ cc_variant + (1|depth)"
      )
    ),
    has_output_dataset=tuple(df1, "Results of LMM with MWD as response and CC variant as predictor variable")
  )
}
```

instance.serialize to file("article.contribution.1.json", format="json-ld")
FAIR scientific information. It’s time.