Proof of concept

Table of contents

The downloaded binary packages are in  
 /var/folders/2x/\_ynr2g956zlgd037jy\_156bm0000gn/T//Rtmp6VOovI/downloaded\_packages

## Introduction

This paper sets out an illustrated approach to developing an end-to-end approach to indicator development, publication and interpretation to assess health status, help develop health policy, and evaluate performance for the Saudi Public Health Authority (SPHA).

SPHA have selected 4 indicators for this proof-of-concept. They are:

1. Percentage of bloodstream infection due to methicillin-resistant Staphylococcus aureus (MRSA)
2. Percentage of bloodstream infection due to Escherichia coli resistant to 3rd-generation cephalosporin
3. Current cigarette smoking among women aged 18-44 years
4. Non-fatal Hospitalizations for All Injuries

### A generic approach

Years of experience suggests a generic approach to developing actionable indicator sets. This is summarised in [Figure 1](#fig-wf).

|  |
| --- |
| Figure 1: Simplified indicator production workflow |

#### The anatomy of an indicator

A key step it so recognise that indicators are numerical values calculated from a numerator and denominator and in a public health context, numerators are event or instance counts and denominators are populations at risk. Public health metrics are generally rates, ratios, or proportions.

Both numerators and denominators will be derived and aggregated from an underlying record system (for example clinical systems, lab systems, vital registration systems, survey databases) which may hold individual records, already aggregated data, or survey results.

The process of generating indicator values will require extracting the right values (business logic, e.g. SQL queries) as part of an Extract-Transform-Load (ETL). Ideally, indicator values would be calculated directly from numerator and denominator values (although this may not always be possible for example if metrics are age-standardised), and should flow and be stored at the **most disaggregated level possible without compromising confidentiality**. This ensures maximum flexibility and responsiveness to policy change, increases statistical power, enables trend analysis and underpins quality assurance.

Therefore a public health indicator system will need a regular flow of highly disaggregated numerator and denominator data from primary data sources. Efficiency dictates that these flows are standarised, and utility that they are as near-real-time as possible.

**Good meta-data** is essential for operating an indicator system and should be linked to each indicator and stored in database format. Each element needs a description of its source, how the value is derived (business logic or calculation details), method of calculation of uncertainty/confidence intervals, as well as rationale, stratifications e.g. temporal disaggregation, age-sex, SES breakdowns and additional information essential for interpretation for example, data discontinuities, recency and so on.

### Illustration #1

#### E.coli cephalosporin resistance

To illustrate I will use English data for E.coli resistance to 3-rd generation cephalosporins. These data are an adjunct to the Public Health Outcome Framework and are publicly available for each NHS hospital in England on a quarterly basis. Detailed metadata is shown in [Table 1](#tbl-amr).

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 1: Detailed metadata   | IndicatorID | definition | description | | --- | --- | --- | | 93188 | Indicator | Rolling quarterly average proportion of 3rd generation cephalosporin resistant E. coli blood specimens; by quarter | | 93188 | Definition | Percentage ofE. coliblood specimens (from English laboratories) with resistant test results fora3rd generation cephalosporin (in the first patient specimen in specimens successfully linked to mandatory surveillanceE. colibacteraemia reports);by acute Trust and by Quarter (2016 Q1 onwards). Presented as a rolling quarterly average across a 4 quarter period.The first quarter presented is Q4 2016, the quarterly average across 2016.3rd generation cephalosporinresistance inE. coliis a key combination highlighted as part of the government 5-year antimicrobial resistance strategy and as such is recommended for monitoring. The testing ofblood specimens for3rd generation cephalosporinsusceptibility is recommended due to the importance of3rd generation cephalosporinswhen treating serious infections.  The 3rd generation cephalosporins considered in this indicator are: cefotaxime, ceftazidime, cefpodoxime and/or ceftriaxone. Tests to one or multiple are taken into account, and the most resistant test result retained.For this indicator, a higher value is indicative of increased3rd generation cephalosporin resistance.<a href="https://www.gov.uk/government/uploads/system/uploads/attachment\_data/file/344071/P\_8i1.1.pdf"> | | 93188 | Rationale | This indicator demonstrates the proportion of patient linked mandatory and laboratory surveillance E. colibacteraemia episodes, presented by reporting acute Trust (mandatory surveillance)that have been tested for susceptibility to3rd generation cephalosporin each NHS acute Trust by quarter, in line with national guidance, and found resistant. A rolling quarterly average (across a 4 quarter period) is presented to comply with data protection regulations when dealing with low numbers.Resistance by acute Trust has been graded by colour in five groups, with the palest colours representing the lowest and the darkest the highest recorded proportions of3rd generation cephalosporin resistance by acute Trust in England.Those where less than 70% linked specimens have been tested in at least one of the last 4 quarters for a particular acute Trust a data quality indicator has been presented alongside the result (70% is the cut off used within the ESPAUR report 2014).If less than 70% of the acute Trust's mandatory E. coli bacteraemia reports have successfully linked to a laboratory report (average across 4 quarters) then a data quality issue is indicated in the presented value. | | 93188 | Data source | Routine voluntary laboratory surveillance reports to UK Health Security Agency via the second generation surveillance system (SGSS); Antimicrobial testing data module | | 93188 | Indicator source |  | | 93188 | Methodology | Thequarterly averagepercentage of resistantantibiotictest results in patient linkedE. coli positive blood specimens recorded by an acute Trust setting in England. Total number of3rd generation cephalosporin resistant test resultson linkedE. colipositive blood specimens for the most recent 4 quarters is aggregated and divided by 4; creating the average quarterly value across the last year. This value is divided by the average quarterly value for the number of3rd generation cephalosporinsusceptibility testson linkedE. colipositive blood specimens across the last yearat the specified geography.Results are multiplied by 100 to be viewed as a percentage. | | 93188 | Standard population/values |  | | 93188 | Confidence interval details | No confidence intervals generated for this indicator | | 93188 | Source of numerator | Voluntary laboratory reports ofEscherichia coli(E. coli) blood specimens made by English laboratories to the routine laboratory surveillance scheme and extracted from UKHSA's Second Generation Surveillance System (SGSS) antimicrobial testing module which have been successfully linked (by patient) to a mandatory surveillance report of E. coli bacteraemia. | | 93188 | Definition of numerator | The quarterly averagenumber(across the last four quarters) of laboratory reports ofEscherichia coli(E. coli) from mandatory surveillance linked blood specimens submitted by English laboratories, tested for and are resistant to a 3rd generation cephalosporin.Data summarises the rolling quarterly average number of reports of3rd generation cephalosporin resistanttest resultsfrom mandatory surveillance linkedEscherichia coli(E. coli)blood specimens submitted to English laboratories.Where multiple blood specimens were taken on the same date these were combined and account for 1 test only, subsequent patient specimens were excluded from the data.The 3rd generation cephalosporins considered in this indicator are: cefotaxime, ceftazidime, cefpodoxime and/or ceftriaxone. Tests to one or multiple are taken into account, and the most resistant test result retained.Data are only available for the NHS and private laboratories signed up to report to UK Health Security Agency’s voluntary surveillance system SGSS.Trust presented is the mandatory surveillance reporting Trust. | | 93188 | Source of denominator | Voluntary laboratory reports ofEscherichia coli(E. coli) blood specimens made by English laboratories to the routine laboratory surveillance scheme and extracted from UKHSA's Second Generation Surveillance System (SGSS) antimicrobial testing module which have been successfully linked (by patient) to a mandatory surveillance report ofE. colibacteraemia. | | 93188 | Definition of denominator | Rolling quarterly average (for the last four quarters) number ofinitial patient laboratory reports ofEscherichia coli(E. coli) identified from blood specimens made by English laboratories to the routine (voluntary) laboratory surveillance scheme and extracted from UKHSA's Second Generation Surveillance System (SGSS) that have been successfully linked to a mandatory surveillance E. coli bacteraemia report, and tested for susceptibility to a 3rd generation cephalosporin.  Data summarise the average number of3rd generation cephalosporin tested reports in linked patientE. coliblood specimens.The 3rd generation cephalosporins considered in this indicator are: cefotaxime, ceftazidime, cefpodoxime and/or ceftriaxone. Tests to one or multiple are taken into account, and considered as one test.  Data are only available for the NHS and private laboratories signed up to report to UK Health Security Agency's voluntary surveillance system SGSS antimicrobial testing module.The presented Trust is the mandatory surveillance reported acute Trust. | | 93188 | Disclosure control |  | | 93188 | Caveats | Data used for this indicator are from the routine voluntary laboratory surveillance scheme following successful linkage (by patient identifier) to the acute Trust level mandatory reports. Total numbers ofE. colibacteraemia presented in this indicator differ from those presented in the mandatory surveillance scheme indicators due to reduced ascertainment in the laboratory reporting scheme.Linkage methodology improvements have been made and used for the Q2 2019 update onwards. Trusts do not report antimicrobial susceptibility information through the mandatory surveillance scheme. The figures produced in this indicator represent a quarterly average proportion of antimicrobial resistance inE. colipositive blood specimen patient episodes; specimens and antimicrobials tested on a subsequent date are not taken into account. Antimicrobial susceptibility test information for multiple patient specimens taken on the same date are combined into a single record with duplicates and repeats being removed. Patient grouping is limited by the entry of key identifiers by reporting laboratories. The data represented reflects specimens reported to UK Health Security Agency's routine laboratory surveillance system antimicrobial testing module (SGSS AMR; formerly AmSurv). Laboratory abilityto report to this system has improved over this time frame, however not all laboratories in England were able to report during the indicator time period. As a result, a zero value does not necessarily reflect a lack of testing in an area but a lack of reporting; interpretation of low values should also take into account the value of the denominator.For the whole time period, values are now presented with only resistant isolates included in the numerator, rather than resistant and/or intermediate (non-susceptible) which was presented previously. This is due to changes in the definition of susceptibility test results used by the European Committee on Antimicrobial Susceptibility Testing (EUCAST), introduced in January 2019. The ‘Intermediate’ susceptibility result has been replaced by a new category ‘Susceptible, increased exposure’ (SIE).The following data quality flags have been applied: Data quality issue due to low reporting in at least one of the last four quarters - where the quarterly average of successfully linked mandatory surveillance E. coli bacteraemia reports for that particular acute Trust is <70%. Possible reasons could be poor patient identifier completion or laboratories within that geographic area being unable to access the laboratory reporting scheme during that time frame.Data quality issue due to low testing in at least one of the last four quarters- where <70% of successfully linked reports were tested for the key antimicrobial (average for the 4 quarter period). Possible reasons could be due to laboratories within that geographic area being unable to access the laboratory reporting scheme during that time frame. The SGSS AMR collection receives reports of all antimicrobial susceptibility tests performed by subscribed laboratories in England. Until April 2016 an irregular suppression mechanism was in place for routine reports from Stoke Mandeville laboratory, applying local antimicrobial stewardship program instructions to laboratory reports prior to their submission to UKHSA, thereby suppressing the fact that the test was performed. As a result there may be under ascertainment of susceptibility testing for specimens from this laboratory; primarily impacting the results in the South Central sub-national region between January 2015 and April 2016. A similar suppression mechanism has been recorded at Barnet Hospital laboratory (part of the Royal Free NHS Trust) and Barts Health microbiology laboratory, resulting in a probable under-ascertainment in the London area. Case data were extracted from the routine laboratory reporting system (SGSS) AMR module on 26th February 2024. Figures reported here may be different to those published elsewhere due to the inclusion of late reports. | | 93188 | Copyright | © Crown copyright | | 93188 | Data re-use |  | | 93188 | Links |  | | 93188 | Indicator number |  | | 93188 | Notes |  | | 93188 | Frequency | Indicator is updated on a quarterly basis | | 93188 | Rounding |  | | 93188 | Indicator Content |  | | 93188 | Specific rationale |  | | 93188 | Simple Name |  | | 93188 | Simple Definition |  | | 93188 | Impact of COVID-19 |  | | 93188 | Unit | % | | 93188 | Value type | Proportion | | 93188 | Year type | Calendar | | 93188 | Polarity | RAG - Low is good | | 93188 | Date updated | 14/03/2024 | |

The metadata provides detail for transparency, and also complies with FAIR principles for reporting (Rothfritz 2019). High quality meta-data is essential for data sharing and reproducibility.

* *Polarity*. This refers to the interpretation and directionality of high (or low) indicator values, specifically if a whether significant value are considered desirable or undesirable i.e. *low is good* or *high is good.* This concept drives the use of colour in data visualisation. For example, low mortality overall, or from any given cause is more desirable than high mortality. Note that mortality is generally compared with a benchmark, and statistically significant difference determined by whether the confidence interval of the indicator contains the benchmark value. If an metric has a definite polarity, it has more use an actionable indicator than a metric without polarity. The PHOF makes extensive use of RAG (red-amber-green) colour palette to denote polarity.
* *Red-red.* If comparisons are made at different scales, of for both temporal and cross-sectional comparisons, a red-red approach is a rapid surveillance tool for detecting strata that are both performing badly, and not making progress. (See example)

Resistance data in England is published quarterly at hospital level and is calculated as rolling year i.e. each point estimate is the annual value to that point in time. This approach smooths the data.

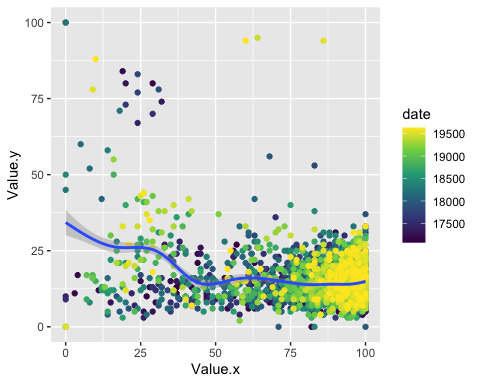
Data is currently available from 2016 Q1 to 2023 Q4 starting with Q4 2016.

#### Trend and variation

|  |  |
| --- | --- |
| |  | | --- | | Figure 2: England Trend in % of E.coli bacteraemias resistant to 3rd generation cephalosporin | |

|  |  |
| --- | --- |
| |  | | --- | | Figure 3: Mean and variation (SD) | |

Data for England suggests that resistance rates have increased as has the variation between hospitals.



All health indicators are measures of performance and performance is only useful through comparison. The level of comparison determines … For example to address the question “How is SA doing internationally” requires comparison with national metric values from comparator countries or regions. To address the question “are we improving?” requires two sets of time series data - the trend for SA for relevant metrics, and the trend for comparator areas or expected trend - which addresses the question “are we improving as expected / fast enough”. The latter question is particularly relevant in relationship to inequalities where the aim might be to reduce the gap between populations.

### Interpreting change

If we find a difference in an indicator value between two points in time, or between two population units, there are three potential explanations:

* The play of chance
* Data artefacts
* Genuine difference

The play of chance can be assessed by calculating uncertainty intervals for our metrics. Data artefacts can be much more difficult to detect and requires understanding of the indicator construction, examining trends in numerators and denominators, and variation for outliers and discontinuities. If differences are not due to chance or data issues, they are likely to reflect a real phenomenon. If differences are real, further investigation may require examining stratified data sets e.g. disaggregation by age or sex or area.

Therefore, to be able to disentangle or identify causality data should be as disaggregated in time, place and person as possible (without compromising confidentiality).

This may not always be possible - for example we may have annual surveys…

For some metrics may take many years to develop meaningful trends

### Policy analysis

In an ideal world we would devise metrics which responded quickly to policy change in order to give rapid feedback on whether our interventions are working. We need to be able to understand the counter-factual - what would have happened without intervention.

| Question | Comparison | Analysis |
| --- | --- | --- |
| What is the problem? |  |  |
| What is the variation? |  |  |
| Are there outliers? |  |  |
|  |  |  |
|  |  |  |

## Analysing regional variation

For this illustration we will simulate a dataset of 50000 bacteraemia observations for 8 regional sub-divisions within Saudi Arabia with variable AMR rates from 3 - 21%.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | area | date | test | id | | --- | --- | --- | --- | | Area6 | 2019-07-01 | No | 1 | | Area2 | 2021-07-01 | No | 2 | | Area7 | 2023-10-01 | No | 3 | | Area3 | 2021-04-01 | No | 4 | | Area7 | 2019-10-01 | Yes | 5 | | Area8 | 2021-04-01 | Yes | 6 | | Area3 | 2022-07-01 | No | 7 | | Area4 | 2022-01-01 | No | 8 | | Area4 | 2021-04-01 | No | 9 | | Area6 | 2022-10-01 | No | 10 | | Area8 | 2019-07-01 | No | 11 | | Area6 | 2020-10-01 | No | 12 | | Area2 | 2020-07-01 | No | 13 | | Area2 | 2023-04-01 | No | 14 | | Area5 | 2020-01-01 | No | 15 | | Area2 | 2023-04-01 | No | 16 | | Area5 | 2019-04-01 | No | 17 | | Area7 | 2022-07-01 | No | 18 | | Area8 | 2022-04-01 | No | 19 | | Area1 | 2021-07-01 | No | 20 |   Figure 4: Simulated data |

|  |
| --- |
| Figure 5: Trend |

|  |
| --- |
| Figure 6: Variation |

Rothfritz, Laura. 2019. “The FAIR Data Principles,” March. <https://doi.org/10.14293/s2199-1006.1.sor-compsci.clnbrup.v1>.