Incidence of confirmed COVID-19 cases in staff and residents in Long Term Care Facilities: Data quality and methodology report

Contents

1 Overview

This report provides additional information on the validation of:

- procedures for classifying the SARS-CoV-2 status (symptomatic, tested, confirmed)
- procedures for computing the at-risk resident-days
- overall numbers of symptomatic and confirmed cases

2 Classification of SARS-CoV-2 cases from incident reports

The SARS-CoV-2 status is determined by a combination of fields from the Datix incidents management system:

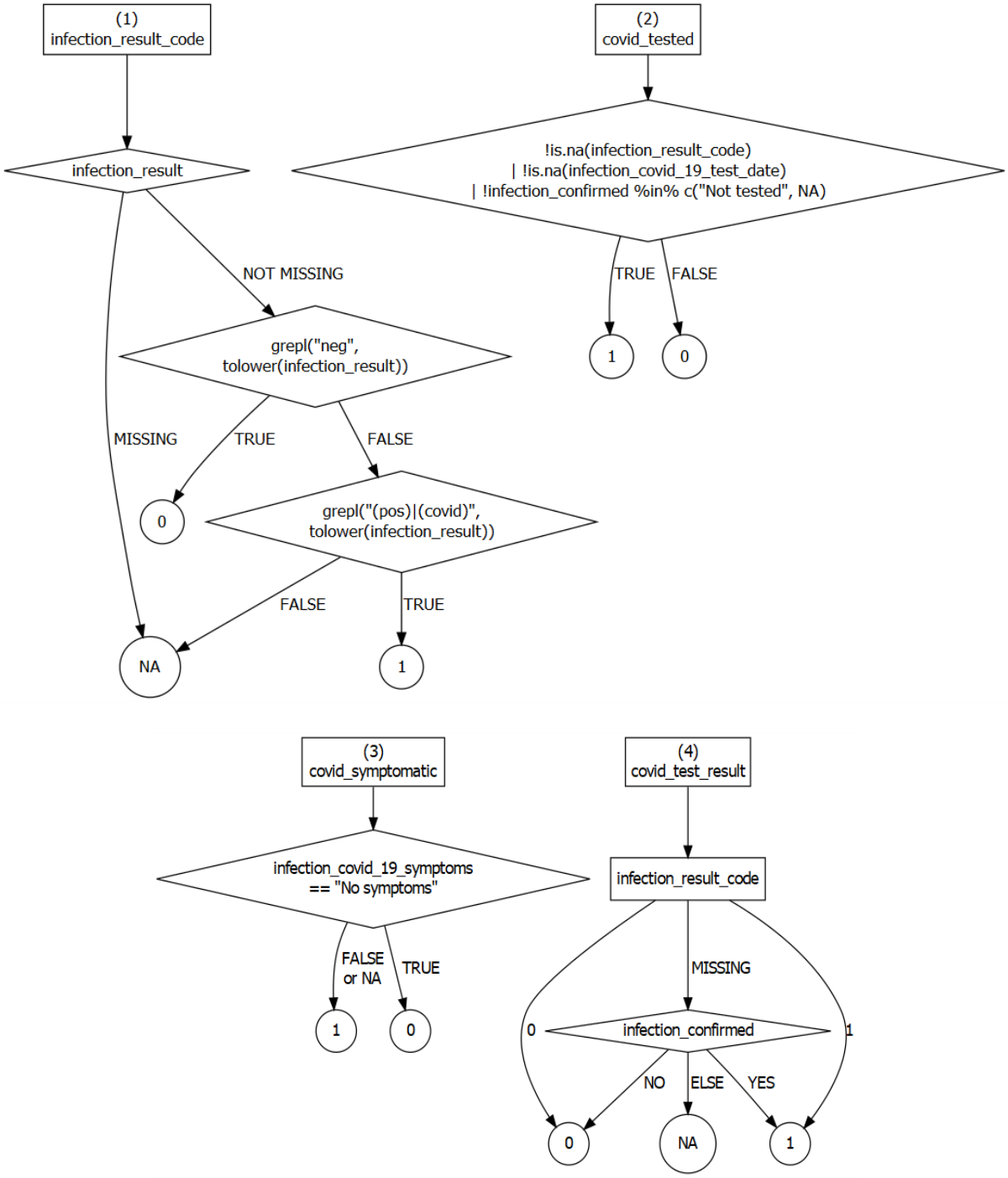
- infection_covid_19_type is a mandatory field one of two possible values:
  - "Confirmed formal clinical diagnosis of COVID-19"
  - "Symptoms, but no definite clinical diagnosis of COVID-19"
- infection_confirmed contains one of 4 possible values (if not missing):
  - "Yes - positive test result: confirmed case"
  - "Test result not yet received"
  - "No - negative test result"
  - "Not tested"
- infection_result is a free-text field containing, for the most part, the words 'positive' or 'negative'

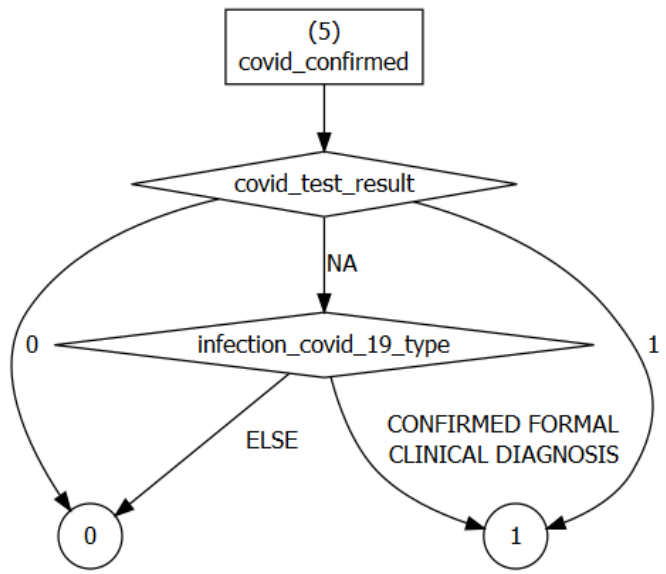
Incident reports are dynamic: they are designed to be updated as more information becomes available. If all fields are not accurately kept updated, some values may become inconsistent. Inconsistencies are observed in incident reports as illustrated in the tabulations below. Due to this, a reclassification algorithm is used which prioritises any information on a positive test result.

	infection_covid_19_type	**Confirmed formal clinical diagnosis of COVID-19**	**Symptoms, but no definite clinical diagnosis of COVID-19**	**<NA>**	**Total**
infection_confirmed					
No - negative test result		10 (4.3%)	224 (95.7%)	0 (0.0%)	234 (100.0%)
Not tested		4 (1.1%)	363 (98.9%)	0 (0.0%)	367 (100.0%)
Test result not yet received		2 (1.7%)	119 (98.3%)	0 (0.0%)	121 (100.0%)
Yes - positive test result: confirmed case		684 (91.9%)	60 (8.1%)	0 (0.0%)	744 (100.0%)
<NA>		0 (0.0%)	13 (100.0%)	0 (0.0%)	13 (100.0%)
Total		700 (47.3%)	779 (52.7%)	0 (0.0%)	1479 (100.0%)

2.1 Algorithm

The following algorithm is used to classify symptomatic, tested, and test-confirmed cases from the Datix reports of COVID-19 infections.





2.2 Validation

Data are tabulated below to evaluate the internal consistency of the rules set out in the previous section.

The infection_result field provides information on test results which broadly agrees yet sometimes conflicts with the infection_confirmed classification of incidents. It is potentially more up to date although it is very sparsely populated.

	covid_confirmed			
infection_covid_19_type	0	1	*Missing*	Total
Confirmed formal clinical diagnosis of COVID-19	11 (1.6%)	689 (98.4%)	0 (0.0%)	700 (100.0%)
Symptoms, but no definite clinical diagnosis of COVID-19	714 (91.7%)	65 (8.3%)	0 (0.0%)	779 (100.0%)
Missing	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Total	725 (49.0%)	754 (51.0%)	0 (0.0%)	1479 (100.0%)

	covid_test_result			
infection_covid_19_type	0	1	*Missing*	Total
Confirmed formal clinical diagnosis of COVID-19	11 (1.6%)	686 (98.0%)	3 (0.4%)	700 (100.0%)
Symptoms, but no definite clinical diagnosis of COVID-19	225 (28.9%)	65 (8.3%)	489 (62.8%)	779 (100.0%)
Missing	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Total	236 (16.0%)	751 (50.8%)	492 (33.3%)	1479 (100.0%)

	covid_test_result			
infection_confirmed	0	1	*Missing*	Total
No - negative test result	234 (100.0%)	0 (0.0%)	0 (0.0%)	234 (100.0%)
Not tested	0 (0.0%)	6 (1.6%)	361 (98.4%)	367 (100.0%)
Test result not yet received	1 (0.8%)	2 (1.7%)	118 (97.5%)	121 (100.0%)
Yes - positive test result: confirmed case	1 (0.1%)	743 (99.9%)	0 (0.0%)	744 (100.0%)
Missing	0 (0.0%)	0 (0.0%)	13 (100.0%)	13 (100.0%)
Total	236 (16.0%)	751 (50.8%)	492 (33.3%)	1479 (100.0%)

3 Data quality

3.1 Data linkage

Secondary use individual-level data for private residents receiving general and elderly care were extracted and pseudonymised by HW. They consisted of two datasets:

- **a residents dataset** containing one record per resident per LTCF, alongside residents' gender, date of birth, most recent dates of admission and discharge, date of first admission, type of stay (residential/nursing) and care (general, dementia, elderly). This dataset does not include any records for local authority 'block contract' beds.
- **an incidents dataset** containing 1,862 reports filed by LTCF staff. Mandated report field were: resident forename, surname, LTCF identifier, incident date/time, and date/time of reporting. In addition, reports could record: date of birth, gender, information on Covid-19 symptoms, tests and test results, resident current location (LTCF/hospital), and death. Incidents were classified to indicate whether the resident was symptomatic, and whether an infection was confirmed on the basis of multiple variables (supplementary material 1).

HW created a resident index file bridging the two datasets using: the first three letters of the resident's forename; the resident's full surname; and the LTCF identifier. Clerical review of identifiers generated in this way established that no single identifier was allocated to distinct individuals in the residents' dataset. Just 64 (3.4%) incident records did not contain enough information to generate an identifier and were excluded. Identifiers of a further 319 records (17.1%) could not be successfully linked to the residents dataset and were excluded, after clerical review established they could not be matched with a resident. A proportion of these records may have been filed in error, while others may relate to occupants of local authority beds who are not accounted for in the residents' dataset.

3.2 Exclusions

A total of 1862 incident reports were obtained. Resident pseudoidentifiers were absent in 64 (3.4%) of Datix incident reports. The pseudoidentifiers of a further 319 reports (17.1%) were could not be successfully linked to the residents database. All were excluded, leaving a total of 1479 reports dated between Monday 2 Mar 2020 and Sunday 7 Jun 2020 relating to 1,197 unique residents. These reports belongued to 126 LTCF's totalling 6,964 beds. For the period of study, the Four Seasons Health Care Group (FSHCG) comprises 179 LTCFs with a total of 9,568 beds.

3.3 Occupancy

LTCF resident-days are available from two sources:

- reports to FSHCG by every LTCF on their occupancy every Sunday, which include residents occupying contract beds
- individual-level resident records of first and last admission and discharge dates, to he exclusion of occupants of contract beds. These records are used to compute occupancy on a daily basis.

In total, 855 out of 9568 beds were contracted by the local authority, with 6 LTCFs having more than half of their beds contracted.

Plots below compare both measures for every Sunday (Figure 1) and in total across the period (Figure 2) at the level of LTCFs, drawing attention to LTCFs with a large proportion of contract beds. The Pearson's correlation between the weekly measures was $r = 0.77$ overall, rising to $r = 0.97$ once excluding LTCFs with $\geq 10\%$ contract beds. These findings confirm that occupancy could be approximated reliably using just residents' records of first and last admission and discharge dates.

This enabled us to estimate exposure (resident-days) excluding local authority contract beds, since no Datix incident reports were returned for occupants of such beds.

In addition, this validation enabled us to approximate the total number of unique residents (including contract beds) during the period of observation for aggregate count data (2020-03-24 – 2020-06-07). The ratio of occupants recorded by FSHCG to the number approximated by us was 1.138. The number of unique residents in non-contract beds was 8,681 during the study period. By assuming that the resident turnover in local authority beds was the same as the rate in other beds, we infer the number of unique residents inclusive of contract beds to be 9,880.

Figure 1: Weekly total resident-days by care home (L=179)

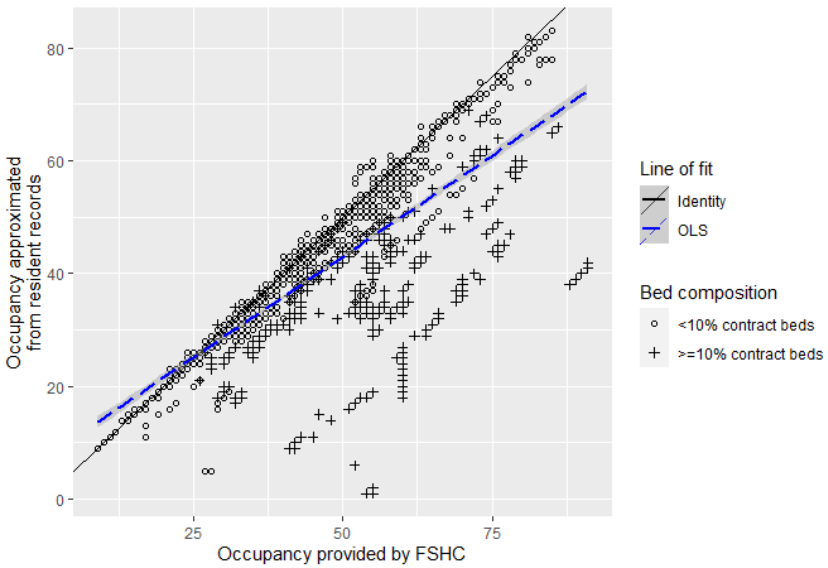
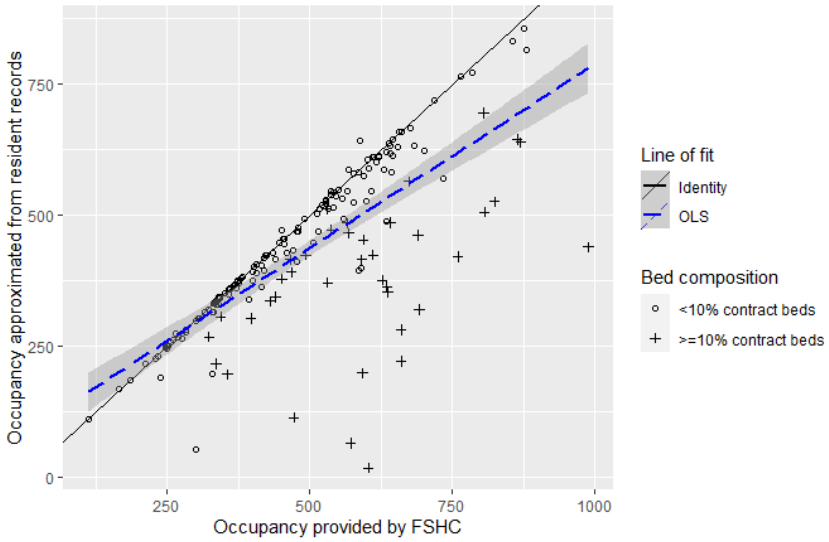


Figure 2: Total resident-days across FSHC during study period (02 Mar 2020 - 07 Jun 2020)



We nevertheless identified 11 LTCFs in which the total resident-days approximated differed by more than 10 percent despite not belonging (Figure 2).

3.4 Comparison of manager-reported tally counts with Datix-derived counts

During the period, manager counted 946 confirmed infection while Datix recorded 601 such infections. The difference could be due to a range of explanations:

1. contract beds occupants, who may represent at least 12.1% of residents (assuming that their turnover is identical to that of occupants of other beds, in reality turnover is likely to be higher, and so is this proportion), may have a higher infection rate than other residents
2. underreporting on Datix
3. linkage error
4. double-counting of cases in manager counts

It is not possible to rely on explanations (1-3) exclusively: the number of confirmed infections on Datix was greater than manager counts in 19 LTCFs. This suggest double counting (4) is a possibility.

Figure 3: Comparison of Datix and manager counts of symptomatic cases

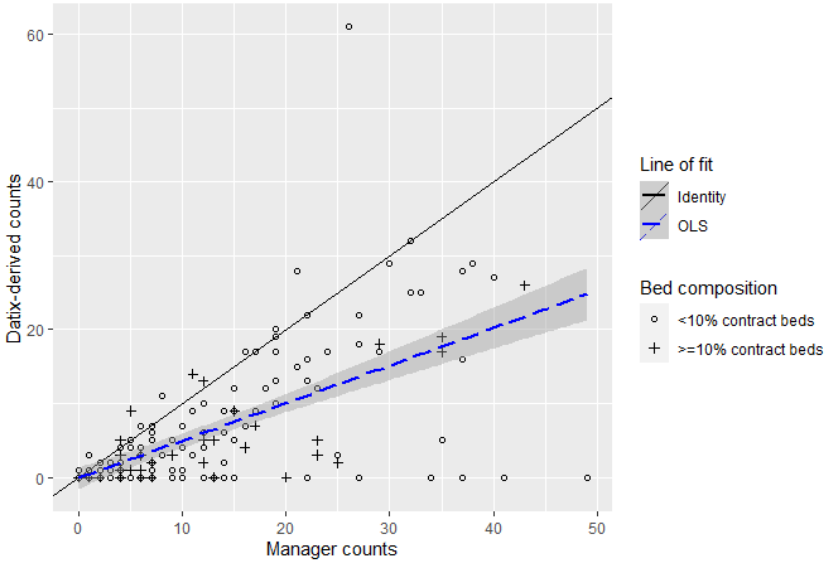
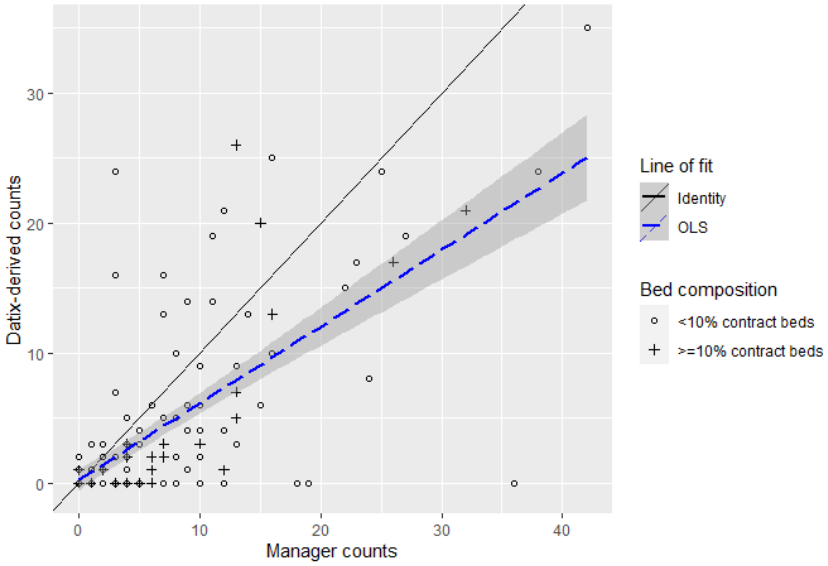


Figure 4: Comparison of Datix and manager counts of confirmed cases



4 Estimating the cumulative incidence of cases and resident-days at risk

Manager-reported outbreak surveillance counts were used together with occupancy estimated from weekly occupied bed censuses in order to estimate the incidence of cases across time using a Kaplan-Meier life table product limit estimator. As the total exposed-to-risk (susceptible resident-days) was unknown, it was approximated in a multiple decrements life table by making the assumption that previously symptomatic/confirmed residents were discharged at the same rate as susceptible residents, as was suggested by the individual-level records. Admitted cases were accounted for in the daily new case count. Based on this $\widehat{I_t}$, the estimated number or previously symptomatic/confirmed residents present in the LTCF on day t was computed as

$$\widehat{I_t} = min(n_t/n_{t-1}, 1)(\hat{E}_{t-1} + d_{t-1}); \qquad I_{t_0} = 0$$

where d_t denotes the number of new cases for each day $t = \{0, \dots, T\}$, while $min(n_t/n_{t-1}, 1)$ acts as an (underestimated) approximation of the discharge rate. The total exposure time was calculated as $\sum_{t=0}^{T}(N_t - \hat{I}_t)$ where N_t denotes daily occupancy figures.

5 Total LTCFs, beds, contract beds, and residents within FSHCG and nationwide

	FSHCG		Nationwide	
	Total LTCFs	**Total beds**	**Total LTCFs**	**Total beds**
England	112	6,085	9,400	45,000
Northern Ireland	43	2,074	483	16,095
Scotland	24	1,409	1,142	40,926

Data sources: (England Care Quality Commission 2020; Northern Ireland Health and Social Care Board 2020; NHS National Services Scotland Information Services Division 2020)

References

England Care Quality Commission. 2020. "Care Directory with Filters." https://www.cqc.org.uk/sites/default/files/HSCA_Active_Locations_01_June_2020.xlsx.

NHS National Services Scotland Information Services Division. 2020. "Care Home Census for Adults in Scotland Figures for 2007-2017 (as at 31 March)." https://www.isdscotland.org/Health-Topics/Health-and-Social-Community-Care/Publications/2018-09-11/2018-09-11-CHCensus-Report.pdf?13142031432.

Northern Ireland Health and Social Care Board. 2020. "Coronavirus (COVID-19) Care Homes - Surge Plan and Frequently Asked Questions." http://www.hscboard.hscni.net/coronavirus/covid-19-care-homes/.