Ethnicity data audit in a secondary care gastroenterology service
Jennifer Randle, Stephen Inns, Ricci Harris, Melissa McLeod

ABSTRACT

AIM: To audit the quality of ethnicity data stored under National Health Index (NHI) in the Hutt Hospital database against a fresh collection of self-identified ethnicity to identify the level of (mis)match present between the datasets.

METHOD: Self-identified ethnicity data was collected from 200 consecutive patients presenting to outpatient gastroenterology services and compared to National Health Index (NHI) in the Hutt Hospital database, using the process outlined in the Primary Care Ethnicity Data Audit Toolkit.

RESULTS: The overall level of match between the individual’s self-identified ethnicity and that recorded in the hospital database was 89% (95% CI [83.8–93.0]). Eighteen patients (9%) self-identified as Māori, 16.7% (95% CI [3.6–41.4]) of whom were not recorded as Māori in the hospital database. Three patients were recorded as Māori in the hospital database but did not self-identify as Māori.

CONCLUSION: Ethnicity data are fundamental to the monitoring and provision of equitable health and healthcare, with a range of applications in the health sector. Our findings of poor-quality ethnicity data for Māori in a hospital NHI database are consistent with previous studies. The assessment of ethnicity data quality must be done in multiple ways to reflect its multiple uses.

Accurate ethnicity data are important because they are used to quantify health inequities, including monitoring trends in disease rates and healthcare delivery. In relation to Māori health equity, high-quality ethnicity data in health are also an obligation under Te Tiriti. In addition, accurate ethnicity data are essential if targeted interventions to address inequity are to be successful, such as the soon to be implemented targeted screening for bowel cancer recommended by a number of Māori health experts. Ethnicity data collection in New Zealand health settings is guided by the Ministry of Health 2017 “HISO 10001:2017 Ethnicity Data Protocols”, which identify the gold standard for collecting ethnicity data in New Zealand as self-reported ethnicity collected using the Stats NZ 2018 Census ethnicity question. These protocols also state that “ethnicity data must be collected at least every three years”. However, ethnicity data in hospital databases are variable in the extent to which they meet these standards.

In New Zealand, ethnicity data have been shown to be less accurate for those identifying as Māori, other minority ethnic groups and for those identifying with multiple ethnicities. There are also variations in the quality of ethnicity data across datasets. There have been several evaluations of New Zealand ethnicity data in a range of contexts over the last two decades. The vast majority have consistently found an undercount of Māori in routinely collected health datasets when they are compared to known high-quality ethnicity data, such as newly collected and standardised self-reported ethnicity. In a 2006 study, a sample of 3,500 patients with diabetes in Waikato found that non-Māori had 99.3% concordance in self-reported ethnicity data (collected for the study) with the hospital database compared to just 71.2% for Māori participants. Eleven years later, a 2017 audit in Waikato District Health Board (WDHB) found Māori undercounts of 17% in the hospital National Health Index (NHI) database and 21% in the trauma registry record. These issues are not only found in hospital data, with a 2007 study of Waitematā Primary Health Organisation (PHO) data showing an undercount for Māori twice that of NZ Europeans when compared to a fresh self-identified collection (on the National Immunisation Collection), and a 2008 study in a large Auckland PHO finding a concordance of 64.9% of Māori, compared with 90.9% of NZ Europeans. Ongoing data quality issues, particularly systematic undercounting of Māori, suggest a level of non-adherence to ethnicity data collection protocols for the health sector.

Undercounts of Māori at the health data aggregate level, and the accuracy of the proportion of Māori (percentage of Māori in the dataset) have important implications. Aggregated data can be used for planning services, measuring rates of disease and in the measurement and monitoring
of health inequities. They are also used for understanding the size of different ethnic groups, and to estimate how many people from different ethnic groups have been reached, such as with vaccination rollouts for Māori. Equally important is the accuracy of ethnicity data at the individual level, where this information may be used to target health services in New Zealand to individuals including the use of ethnicity as part of cardiovascular risk assessment, diabetes screening and sore throat swabbing. Information on the accuracy of ethnicity data at the individual level can also help inform us about the quality of aggregate level ethnicity data, as it provides important information on ethnicity misclassification and whether this is differential or not.

This audit aims to comprehensively evaluate the accuracy of ethnicity data at both the aggregate and individual levels by comparing a fresh collection of self-identified ethnicity data with the ethnicity data stored under NHI in the Hutt Hospital database for a sample of 200 consecutive patients presenting to Hutt DHB gastroenterology services. Hutt Hospital is an urban tertiary teaching centre, where the gastroenterology department cared for more than 13,000 patients who attended outpatient clinics and endoscopies between 1 April 2020 and 31 March 2021. It provides care to those living in the Hutt Valley District Health Board (HVDHB) region, serving a population of 156,790 people—17.9% of whom are estimated to be Māori based on 2020/2021 population projections by the Ministry of Health.

Methods

This study was a prospective ethnicity audit of 200 consecutive patients aged 16 years and over presenting to the HVDHB gastroenterology service as outpatients to either a gastroenterology clinic or an endoscopy between 2–11 November 2021. The aim of this study was to audit the quality of ethnicity data by comparing current self-identified ethnicity (directly collected from patients for this audit), hereby referred to as “self-identified ethnicity”, to that recorded under the NHI in the Hutt Hospital database. We identified the level of match or mismatch present between the collections. We note that white ethnicity in all health data is also supposed to be self-identified according to protocols for the health sector, that is not always the case.

Data collection and analysis

This audit followed the methods described in the Ministry of Health’s “Primary Care Ethnicity Data Toolkit” published in 2013. Ethnicity data were collected from consecutive patients on each day of data collection until a total sample size of 200 patients was achieved. The protocol methods align with the Ministry of Health 2017 “HISO 10001:2017 Ethnicity Data Protocols” and thus self-identified ethnicity collected in this study is considered the “gold standard” in comparison to the NHI ethnicity recorded on the hospital NHI database.

The audit was powered to detect whether a mismatch is present between self-identified ethnicity data and the NHI hospital database ethnicity data. With 200 patients total and 90% accuracy in ethnicity data overall, the overall margin of error is roughly +/-5% (proportion/accuracy=90, lower/upper bound of 95% CI, 85.0/93.8).

All self-reported ethnicity data were collected by JR using the 2001 New Zealand Census ethnicity question via a patient-completed form after verbal consent was obtained. From the time of invitation to take part in the audit until the patient had completed their self-identified ethnicity, the data-collector was blind to the hospital database ethnicity data. Patients’ full names and dates of birth were also collected for the purposes of correctly identifying the patients in the hospital database. NHI-linked ethnicity data in the HVDHB Hospital database for each patient were then accessed and recorded by JR. All hospital database ethnicity data were collected on the same day as the patient presentation. Ethnicity data from both sources were coded to Level 2 of the 2004 Ministry of Health ethnicity data protocols for analysis.

We used three measures to evaluate the accuracy of the hospital NHI database: ethnic proportions, individual mismatch and undercounts. Ethnic proportions were calculated separately for the self-identified collection and the hospital database (NHI) ethnicity, and were calculated as the number of individuals who reported being a member of the ethnic group divided by the total population size in the health dataset. For each individual, the level of match of ethnicity data between self-identified and NHI hospital database was categorised as match, partial match and total mismatch (Table 1). The measurement of undercount assumes that self-identified ethnicity is the gold standard collection. It gives the proportion of individuals who self-identified an ethnic group that were not recorded with that ethnic group in the hospital database.
**Statistical analysis**

Data were analysed in Microsoft Excel version 2110 and R Studio (Version 1.4.1106) to generate descriptive statistics and confidence intervals.

**Results**

Of the 202 patients invited to take part in the study, 200 (99%) consented to participate. The median age of participants was 56.5 years (IQR 42–69.25) with an age range of participants from 16 to 93 years. Slightly more participants were female (n=107, 53.5%). Participants were a mix of those presenting for endoscopy (n=114, 57%) and gastroenterology clinics (n=86, 43%).

In the sample of 200 patients, 77.5% (n=155) self-identified as New Zealand European (either alone or in combination with other ethnicities) and 9% (n=18) identified themselves as Māori (Table 2). Half of self-identified Māori reported belonging to more than one ethnic group (9/18) compared to just 3% (5/182) of non-Māori (Table 3). Overall, those self-identifying with multiple ethnicities had a 50% (95% CI, 23.0–77.0) match (with all ethnicities also recorded on the hospital NHI database) compared to 91.9% (95% CI, 87.0–95.4) for patients who self-identified with a single ethnicity (Table 3). All mismatches in the multiple ethnicities group were partial mismatches, where some but not all of the self-identified ethnicities were recorded.

The overall level of match between the self-identified and the hospital NHI database ethnicity was 89% (95% CI, 83.8–93.0). Partial mismatches were found for 5% (10/200) of patients and total mismatch for 6% (12/200). Among non-Māori patients only 3.8% (95% CI, 1.6–7.8) of patients had partial mismatch of ethnicity data and 6.6% (95% CI, 3.5–11.2) had total mismatch of data. In comparison the match for Māori was 83.3% (95% CI, 58.6–96.4), 16.7% (95% CI, 3.6–41.4) of Māori patients had partial mismatch of ethnicity data and there were no total mismatches. Importantly, all the partial mismatches for Māori resulted in the patient being unable to be identified as Māori on the hospital NHI database. This means that the undercount for Māori on the hospital NHI database when compared to the self-identified fresh collection was 16.7%. The undercount in the NZ European group was 5.8%.

In both the self-reported ethnicity data and the hospital NHI database, the ethnic proportion of Māori was 9% (18/200). There were three patients who were identified as Māori in the hospital NHI database that did not self-report Māori ethnicity—two of whom identified themselves as NZ European only, and one who self-identified as NZ European and Niuean. Additionally, there were three patients who were identified as NZ European only in the hospital NHI database but self-identified as both Māori and NZ European.

There was some variation in the level of match for ethnicity when stratified by other demographics. Males had a lower match rate (81.7%) than females (95.3%), with a higher rate of total mismatch in males (11.8%) compared to females (0.9%). Most participants were aged between 40 and 79, and mismatch was relatively evenly distributed by age of around 10–15% mismatch, with the exception of being slightly higher in the 70–79-year-old category at 17% (Figure 1). The proportion of (mis)match did not differ between patients presenting to clinic or endoscopy.

**Discussion**

This study audited the accuracy of NHI ethnicity data in a hospital NHI database against self-reported ethnicity, using the methods described in the Ministry of Health’s “Primary Care Ethnicity Data Toolkit”. We found an overall match rate of 89% in our sample and a match rate of 83.3% for Māori.

All the mismatches found for people self-reporting Māori ethnicity in our sample resulted in them being unidentifiable in the hospital NHI database, with an undercount of Māori on the hospital NHI database measuring 16.7%. In comparison, the undercount for the NZ European group was 5.8%. Undercounts of Māori have been found in several studies over the last two decades in a variety of areas of the health system. Swan et al. audited the accuracy of ethnicity data in a group of 3,500 diabetes patients in the Waikato in 2006 using a non-standard ethnicity question, and in 2007 Bramley et al. looked at 57,612 vaccination records of school-aged children with data collected using the census ethnicity question. These studies found Māori undercounts of 28.8% and 37.1% respectively. Undercounts found in more recent studies have been more variable. A 2017 study by Scott et al. found an undercount of Māori by 17% in the hospital NHI database and 21% in the Trauma Registry. In contrast, another 2017 study of 227 patients with kidney disease found a Māori undercount of 7.3%, but this was still twice the undercount of 3.6% found in NZ Europeans. The results of our audit would support evidence
Table 1: Coding classifications for level of match of ethnicity data.

<table>
<thead>
<tr>
<th>Level of match</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Match (M)</td>
<td>All ethnicities identified by the patient on the audit form match with those recorded for that patient in the database. The order in which they are recorded does not need to match.</td>
</tr>
<tr>
<td>Partial match (PM)</td>
<td>This is only relevant for patients with multiple ethnicities recorded and applies where some, but not all, of the ethnicities match, regardless of the order.</td>
</tr>
<tr>
<td>Total mismatch (TMM)</td>
<td>None of the ethnicities identified by the patient on the audit form match with those recorded for that patient in the Patient Management System (PMS).</td>
</tr>
</tbody>
</table>

Table 2: Overall ethnic group proportions by data source using total response ethnicity.

<table>
<thead>
<tr>
<th>Overall ethnicity count</th>
<th>Hospital NHI database</th>
<th>%</th>
<th>Self-reported</th>
<th>%</th>
<th>Change in group size</th>
</tr>
</thead>
<tbody>
<tr>
<td>NZ European</td>
<td>148</td>
<td>74.0%</td>
<td>155</td>
<td>77.5%</td>
<td>7</td>
</tr>
<tr>
<td>Māori</td>
<td>18</td>
<td>9.0%</td>
<td>18</td>
<td>9.0%</td>
<td>0</td>
</tr>
<tr>
<td>Other European</td>
<td>14</td>
<td>7.0%</td>
<td>10</td>
<td>5.0%</td>
<td>-4</td>
</tr>
<tr>
<td>Chinese</td>
<td>9</td>
<td>4.5%</td>
<td>9</td>
<td>4.5%</td>
<td>0</td>
</tr>
<tr>
<td>SE Asian</td>
<td>7</td>
<td>3.5%</td>
<td>8</td>
<td>4.0%</td>
<td>1</td>
</tr>
<tr>
<td>Indian</td>
<td>7</td>
<td>3.5%</td>
<td>7</td>
<td>3.5%</td>
<td>0</td>
</tr>
<tr>
<td>Niuean</td>
<td>0</td>
<td>0.0%</td>
<td>1</td>
<td>0.5%</td>
<td>1</td>
</tr>
<tr>
<td>Samoan</td>
<td>1</td>
<td>0.5%</td>
<td>1</td>
<td>0.5%</td>
<td>0</td>
</tr>
<tr>
<td>Other Asian</td>
<td>1</td>
<td>0.5%</td>
<td>1</td>
<td>0.5%</td>
<td>0</td>
</tr>
<tr>
<td>Fijian</td>
<td>1</td>
<td>0.5%</td>
<td>1</td>
<td>0.5%</td>
<td>0</td>
</tr>
<tr>
<td>Cook Island Māori</td>
<td>1</td>
<td>0.5%</td>
<td>1</td>
<td>0.5%</td>
<td>0</td>
</tr>
<tr>
<td>African</td>
<td>1</td>
<td>0.5%</td>
<td>1</td>
<td>0.5%</td>
<td>0</td>
</tr>
<tr>
<td>Response Unidentifiable</td>
<td>0</td>
<td>0.0%</td>
<td>1</td>
<td>0.5%</td>
<td>1</td>
</tr>
</tbody>
</table>
Table 3: Proportion of patients who self-identified with more than one ethnic group using total response ethnicity.

<table>
<thead>
<tr>
<th>Ethnic group</th>
<th>Total (n=214)*</th>
<th>More than one ethnic group reported</th>
<th>% of ethnic group reporting multiple ethnicities</th>
</tr>
</thead>
<tbody>
<tr>
<td>NZ European</td>
<td>155</td>
<td>13</td>
<td>8.4%</td>
</tr>
<tr>
<td>Māori</td>
<td>18</td>
<td>9</td>
<td>50.0%</td>
</tr>
<tr>
<td>Other European</td>
<td>10</td>
<td>2</td>
<td>20.0%</td>
</tr>
<tr>
<td>Chinese</td>
<td>9</td>
<td>1</td>
<td>11.1%</td>
</tr>
<tr>
<td>SE Asian</td>
<td>8</td>
<td>1</td>
<td>12.5%</td>
</tr>
<tr>
<td>Indian</td>
<td>7</td>
<td>1</td>
<td>14.3%</td>
</tr>
<tr>
<td>Niuean</td>
<td>1</td>
<td>1</td>
<td>100.0%</td>
</tr>
<tr>
<td>Samoan</td>
<td>1</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>Other Asian</td>
<td>1</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>Fijian</td>
<td>1</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>Cook Island Māori</td>
<td>1</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>African</td>
<td>1</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>Response Unidentifiable</td>
<td>1</td>
<td>0</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

*Individuals were included in every ethnic group that they identified with (a total response approach), meaning that some individuals are counted in multiple ethnic groups.

Figure 1: Any mismatch by age for total sample population.
that there remains inequity between the size of the undercounts for Māori when compared to the NZ European population.

In our study, the ethnic proportions of Māori in the self-reported ethnicity data and the hospital NHI database were both 9%. This finding differs from other New Zealand-based hospital ethnicity audits that have found both an undercount of Māori participants in the hospital NHI database and a lower ethnic proportion of Māori. This difference may reflect the smaller sample of Māori in our audit due to the lower ethnic proportion of Māori in the HVDHB catchment and potential selection bias in terms of who is able to navigate the healthcare pathway to access gastroenterology services (discussed further below).

At the individual level, we found differences in the individuals identified/recorded as Māori in the datasets, with three participants identified as Māori in the hospital NHI database who did not self-report Māori ethnicity, and three participants not identified as Māori in the hospital NHI database who did self-identify as Māori. This has important implications for our ability to identify the correct individuals to invite to existing and future health programmes that use NHI ethnicity data to target individuals. Both the cardiovascular disease risk assessment and diabetes screening in primary care start for Māori, Pasifika and South Asian patients at an earlier age. Being of Māori or Pacific ethnicity is also used in guidance for sore throat swabbing for group A streptococcal pharyngitis to prevent rheumatic fever. More recently, COVID-19 vaccination efforts for Māori relied on the accurate identification of individuals (as Māori) from existing data collections, as well as the use of aggregate-level denominators sourced from health service data, for reporting of vaccine coverage.

Similar to other studies, we found that there was lower accuracy for those self-identifying with multiple ethnicities. A complete match was found for only 50% of the 14 participants that had multiple self-reported ethnicities, compared to 91.9% in those who only self-reported one ethnicity. Importantly, half of the Māori participants (compared to just 3% of non-Māori) in our sample self-reported multiple ethnicities and so the impact of this inaccuracy may be greater for Māori. Consistent with the Primary Care Ethnicity Data Audit Toolkit our audit was powered to detect overall mismatch and was therefore underpowered to detect differences for multiple ethnicities. In addition, due to a smaller Māori sample than anticipated, some of our confidence intervals around our best estimates are overlapping. However, our findings are consistent with other research and indicate an ongoing issue with ethnicity data quality.

Māori were under-represented in our sample, making up just 9% of patients, compared to 17.9% in the HVDHB population (a difference that remained when we limited the sample to those aged 30–69 years at 10% versus 16.7% for HVDHB). While there is the possibility of sampling error (in terms of Māori just not being booked on the days the audit was run), it is likely that the under-representation of Māori in our sample also reflects that Māori have inequitable access to this service. In general, Māori carry a higher rate of comorbidities and have higher rates of a large number of gastroenterological illnesses including stomach cancer, hepatocellular carcinoma, gastric ulcers and upper gastrointestinal bleeding, and similar rates of bowel cancer (but worse survival). There are lower rates of diagnosed inflammatory bowel disease in Māori. On balance, we expect Māori have greater need for gastroenterological services than non-Māori, and so we were surprised to find that in this study Māori were underrepresented in the gastroenterology service.

The inequitable rates of gastroenterology service use identified in this study may be an indication of barriers along the healthcare pathway that are experienced by Māori. There are well-documented inequities in access to primary care, and a higher level of unmet healthcare need for Māori, which then impacts on access to secondary care. Those able to navigate the healthcare system into secondary care may have had more opportunities to update their ethnicity data. For example, all patients presenting to a gastroenterology clinic or endoscopy at Hutt Hospital are asked to review a paper copy of their hospital details at each presentation, including their ethnicity information. Our sample of patients, particularly those who saw gastroenterology services regularly, would have had multiple opportunities to review and change their ethnicity information, potentially (although not necessarily) improving accuracy. Therefore, our estimate of the level of undercount of Māori ethnicity may be smaller than the undercount for those less able to navigate the healthcare system, in settings where patients can present to services directly (e.g., emergency department), or in other district health boards with different processes around ethnicity data collection, and recording.
Ethnicity data collection needs to be an ongoing activity as ethnicity can change over time, and this has been shown to be a greater issue for Māori, Pasifika and Asian populations more than other ethnicities. A 2009 New Zealand study found that 8% of individuals changed their self-reported ethnicity over a 3-year period, and the strongest predictors of this change were reporting Māori, Pasifika or Asian identity at the beginning of the study, as well as reporting multiple ethnic groups. This also means that the ability to regularly update self-reported ethnicity will have an effect on accuracy, and that this will impact Māori and other minoritised groups more than NZ Europeans.

All areas of health in New Zealand need to collect, and maintain access to, high-quality ethnicity data. The New Zealand Ministry of Health ethnicity data protocols published in 2017 already provide guidance on ways to improve ethnicity data quality. This guidance includes ensuring ethnicity is self-identified, using only the census ethnicity question for collection and a maximum of 3 years between each collection of ethnicity data. However, this guidance requires supported implementation to be followed appropriately with ongoing quality assessment. There are also other opportunities to collect updated ethnicity data that could improve quality on a national level that could form part of a wider strategy, such as collection of updated data as part of vaccination programmes. A key finding in this study is the importance of measuring the accuracy of ethnicity data in multiple ways to reflect their multiple uses.

Conclusion

Ethnicity data are fundamental to the monitoring and provision of equitable health and healthcare, with a range of applications in the health sector. Data can accurately reflect ethnic proportions at a group level but be inaccurate at the individual level, which can be falsely reassuring about overall data quality. Our findings of poor-quality ethnicity data in the hospital NHI record are consistent with previous studies. Māori have been shown to be particularly affected by inaccurate ethnicity data in a range of settings that are important for measuring and acting to correct health inequities.
COMPETING INTERESTS
The authors have no conflicts of interest or financial disclosures.

ACKNOWLEDGEMENTS
The authors would like to thank the study participants and the staff from the gastroenterology services unit in Hutt Hospital and the research office at CCDHB/HVDHB for their local approval and support. Ethical approval for the study was obtained from the University of Otago Human Ethics Committee (Health) (Reference HD21/080), the HVDHB Clinical Audit Committee (Reference: 2021/44) and the Ngāi Tahu Research Consultation Committee (Reference: 5926_22803).

AUTHOR INFORMATION
Jennifer Randle: Public Health Medicine Registrar, Department of Public Health, University of Otago, Wellington, New Zealand.

Stephen Inns: Gastroenterologist, Hutt Valley District Health Board, Lower Hutt. Senior Lecturer, Department of Medicine, University of Otago, Wellington, New Zealand.

Ricci Harris: Public Health Physician/Research Associate Professor, Department of Public Health, University of Otago, Wellington, New Zealand.

Melissa McLeod: Public Health Physician/Senior Lecturer, Department of Public Health, University of Otago Wellington, New Zealand.

CORRESPONDING AUTHOR
Jennifer Randle: Public Health Medicine Registrar, Department of Public Health, University of Otago, Wellington, New Zealand. E: jenrandle@live.com

REFERENCES
16. Ministry of Health. Primary Care Ethnicity


