CHAPTER TWO
Methodology
CHAPTER 2: METHODOLOGY

The 2014 National Medical Care Survey (NMCS) is a national cross-sectional study of primary care activities. It utilised a multi-stage stratified cluster sampling design, with the primary care clinics acting as primary sampling units (PSUs). Random sampling of primary care clinics was performed for all states and federal territories in Malaysia, namely, Johor, Kedah, Kelantan, Melaka, Negeri Sembilan, Pahang, Perak, Perlis, Pulau Pinang, Sabah, Sarawak, Selangor, Terengganu and Wilayah Persekutuan (WP) Kuala Lumpur. Two federal territories were combined with the neighbouring states (WP Labuan with Sabah and WP Putrajaya with Selangor) in view of the geographical proximity and demographic similarities.

The data collection lasted 17 weeks, from 7 January 2014 to 15 May 2014. All sampled clinics were randomly allocated one day for data recording in their respective clinics, and all service providers working on that particular day were involved in data collection.

2.1 SAMPLE SIZE CALCULATION AND SAMPLING METHODS

Ideally, we would like to randomly sample the units of analysis which are the encounters; however this is not feasible in our current system. The reasons being we do not have an exhaustive list of primary care patients and it would not be practical, financially and logistically, to sample patients from all over the country. Hence the sampling could only be done via the clinics which act as a cluster of encounters. The cluster effect of such sampling method will be adjusted in the analysis using statistical programme.

Sample size calculation

The number of encounters needed for the NMCS 2014 was first determined for each sector based on the formula proposed by Cochran\(^1\) by using the proportion of upper respiratory tract infection encounters from NMCS 2010. This number was then adjusted for the design effect (assumed to be 2) and expected response rate from each sector.

Subsequently, the adjusted number of encounters was proportionately distributed to each state and by using the average number of doctors per clinic and the average number of encounters per doctor from NHEWS Primary Care 2010, the number of clinics to be sampled for each stratum was calculated. We expected a minimum of 30 encounters from each clinic.

The final sample consisted of 139 public clinics and 1,002 private clinics (Table 2.1.1). For Melaka and Perlis, all public clinics were sampled because the total number of clinics in these strata was less than 30, the minimum acceptable sample size for each stratum.\(^2\)
The 2014 National Medical Care Survey (NMCS) is a national cross-sectional study of primary care activities. It utilised a multi-stage stratified cluster sampling design, with the primary care clinics acting as primary sampling units (PSUs). Random sampling of primary care clinics was performed for all states and federal territories in Malaysia, namely, Johor, Kedah, Kelantan, Melaka, Negeri Sembilan, Pahang, Perak, Perlis, Pulau Pinang, Sabah, Sarawak, Selangor, Terengganu and Wilayah Persekutuan (WP Kuala Lumpur). Two federal territories were combined with the neighbouring states (WP Labuan with Sabah and WP Putrajaya with Selangor) in view of the geographical proximity and demographic similarities.

The data collection lasted 17 weeks, from 7 January 2014 to 15 May 2014. All sampled clinics were randomly allocated one day for data recording in their respective clinics, and all service providers working on that particular day were involved in data collection.

2.1 SAMPLE SIZE CALCULATION AND SAMPLING METHODS

Ideally, we would like to randomly sample the units of analysis which are the encounters; however this is not feasible in our current system. The reasons being we do not have an exhaustive list of primary care patients and it would not be practical, financially and logistically, to sample patients from all over the country. Hence the sampling could only be done via the clinics which act as a cluster of encounters. The cluster effect of such sampling method will be adjusted in the analysis using statistical programme.

Sample size calculation

The number of encounters needed for the NMCS 2014 was first determined for each sector based on the formula proposed by Cochran 1 by using the proportion of upper respiratory tract infection encounters from NMCS 2010. This number was then adjusted for the design effect (assumed to be 2) and expected response rate from each sector.

Subsequently, the adjusted number of encounters was proportionately distributed to each state and by using the average number of doctors per clinic and the average number of encounters per doctor from NHEWS Primary Care 2010, the number of clinics to be sampled for each stratum was calculated. We expected a minimum of 30 encounters from each clinic.

The final sample consisted of 139 public clinics and 1,002 private clinics (Table 2.1.1).

<table>
<thead>
<tr>
<th>State/federal territory</th>
<th>Public Population</th>
<th>Public Sample</th>
<th>Private Population</th>
<th>Private Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Johor</td>
<td>93</td>
<td>11</td>
<td>709</td>
<td>117</td>
</tr>
<tr>
<td>Kedah</td>
<td>56</td>
<td>7</td>
<td>298</td>
<td>55</td>
</tr>
<tr>
<td>Kelantan</td>
<td>64</td>
<td>7</td>
<td>192</td>
<td>45</td>
</tr>
<tr>
<td>Melaka</td>
<td>29</td>
<td>26</td>
<td>186</td>
<td>31</td>
</tr>
<tr>
<td>Negeri Sembilan</td>
<td>46</td>
<td>6</td>
<td>233</td>
<td>40</td>
</tr>
<tr>
<td>Pahang</td>
<td>79</td>
<td>6</td>
<td>201</td>
<td>42</td>
</tr>
<tr>
<td>Perak</td>
<td>83</td>
<td>11</td>
<td>510</td>
<td>73</td>
</tr>
<tr>
<td>Perlis</td>
<td>9</td>
<td>9</td>
<td>30</td>
<td>10</td>
</tr>
<tr>
<td>Pulau Pinang</td>
<td>30</td>
<td>5</td>
<td>398</td>
<td>70</td>
</tr>
<tr>
<td>Sabah &amp; WP Labuan</td>
<td>92</td>
<td>9</td>
<td>311</td>
<td>65</td>
</tr>
<tr>
<td>Sarawak</td>
<td>196</td>
<td>10</td>
<td>225</td>
<td>34</td>
</tr>
<tr>
<td>Selangor &amp; WP Putrajaya</td>
<td>76</td>
<td>15</td>
<td>1,520</td>
<td>270</td>
</tr>
<tr>
<td>Terengganu</td>
<td>45</td>
<td>5</td>
<td>148</td>
<td>30</td>
</tr>
<tr>
<td>WP Kuala Lumpur</td>
<td>13</td>
<td>12</td>
<td>685</td>
<td>120</td>
</tr>
<tr>
<td>Total</td>
<td>911</td>
<td>139</td>
<td>5,646</td>
<td>1,002</td>
</tr>
</tbody>
</table>

Sampling methods

The sampling frame of public and private clinics was generated by matching the list of clinics from National Healthcare Establishments and Workforce Survey (NHEWS) 2012 with several sources:

- The list of public clinics (Klinik Kesihatan) from the Family Health Development Division, Ministry of Health (MOH) Malaysia.
- The list of registered private clinics from the Private Medical Practice Division, Ministry of Health Malaysia (often referred to as the Cawangan Kawalan Amalan Perubatan Swasta (CKAPS)).

Both lists were updated as of 31st December 2012 and these were regarded as the most recent lists of the public and private clinics at the period of survey.

As for clinics that were not matched from the lists, subsequent verification by telephone calls was done to determine the existence or current operational status of the establishments. Those that were found to be closed or do not meet our inclusion and exclusion criteria were removed from the sampling frame (Table 2.1.2).
Table 2.1.2: Inclusion and exclusion criteria for the clinics sampled in the survey

<table>
<thead>
<tr>
<th>Inclusion criteria</th>
<th>Exclusion criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>• MOH Health Clinics <em>(Klinik Kesihatan)</em> which provide primary care services</td>
<td>• Outpatient departments within hospital or maternity homes</td>
</tr>
<tr>
<td>• Private medical clinics registered with CKAPS and provide primary care services</td>
<td>• Public clinics with the following criteria:</td>
</tr>
<tr>
<td></td>
<td>– Health clinics without permanent medical doctors <em>(Klinik Kesihatan)</em></td>
</tr>
<tr>
<td></td>
<td>– Clinics which provide only maternal and child health services <em>(Klinik Kesihatan Ibu dan Anak)</em></td>
</tr>
<tr>
<td></td>
<td>– Rural health clinics <em>(Klinik Desa)</em></td>
</tr>
<tr>
<td></td>
<td>– 1 Malaysia clinics</td>
</tr>
<tr>
<td></td>
<td>• Private clinics with the following criteria:</td>
</tr>
<tr>
<td></td>
<td>– Aesthetic clinics</td>
</tr>
<tr>
<td></td>
<td>– Charity clinics</td>
</tr>
<tr>
<td></td>
<td>– Diagnostic centres</td>
</tr>
<tr>
<td></td>
<td>– Homeopathy clinics</td>
</tr>
<tr>
<td></td>
<td>– In-house clinics/clinics which are affiliated with specific companies</td>
</tr>
<tr>
<td></td>
<td>– Specialist clinics /clinics which provide specialised care/ e.g. paediatric, cardiology, occupational therapy</td>
</tr>
<tr>
<td></td>
<td>– Clinics which operate less than 5 days a week</td>
</tr>
<tr>
<td></td>
<td>– Clinics which participated in NMCS 2012</td>
</tr>
</tbody>
</table>

Sample selection was conducted by stratified random cluster sampling, incorporating several stages. The details are described below.

*Stratification*

**Stage 1: Stratification by sector**

– Each state or federal territory was stratified by either public or private sector.

**Stage 2: Stratification by sampling regions**


*Cluster sampling*

**Stage 1: Sampling of clinics (primary sampling unit)**

– Random sampling of clinics was based on random numbers generated using Microsoft Excel 2007.
– If a selected clinic was discovered to not fulfill the inclusion and exclusion criteria when contacted, the clinic was omitted and another clinic was randomly selected to replace it.
Chapter 2: Methodology

Table 2.1.2: Inclusion and exclusion criteria for the clinics sampled in the survey

**Inclusion criteria**
- MOH Health Clinics (Klinik Kesihatan) which provide primary care services
- Private medical clinics registered with CKAPS and provide primary care services

**Exclusion criteria**
- Outpatient departments within hospital or maternity homes
- Public clinics with the following criteria:
  - Health clinics without permanent medical doctors (Klinik Kesihatan)
  - Clinics which provide only maternal and child health services (Klinik Kesihatan Ibu dan Anak)
  - Rural health clinics (Klinik Desa)
  - 1 Malaysia clinics
- Private clinics with the following criteria:
  - Aesthetic clinics
  - Charity clinics
  - Diagnostic centres
  - Homeopathy clinics
  - In-house clinics/clinics which are affiliated with specific companies
  - Specialist clinics /clinics which provide specialised care/ e.g. paediatric, cardiology, occupational therapy
  - Clinics which operate less than 5 days a week
  - Clinics which participated in NMCS 2012

Sample selection was conducted by stratified random cluster sampling, incorporating several stages. The details are described below.

**Stage 1: Stratification by sector**
- Each state or federal territory was stratified by either public or private sector.

**Stage 2: Stratification by sampling regions**

**Cluster sampling**
- **Stage 1: Sampling of clinics (primary sampling unit)**
  - Random sampling of clinics was based on random numbers generated using Microsoft Excel 2007.
  - If a selected clinic was discovered to not fulfill the inclusion and exclusion criteria when contacted, the clinic was omitted and another clinic was randomly selected to replace it.

**Stage 2: Sampling of survey date (secondary sampling unit)**
- Each sampled clinic was randomly assigned a date for data collection within the study period.
- The following days were excluded:
  - Public holidays
  - Weekends, including Friday, Saturday and Sunday
  - Monday (Mondays are usually the busiest for public primary care clinics)
  - A week before and during the festive season (Chinese New Year)
- If the clinic was closed on the day of survey, the doctor had the option to change the survey date to the next available working day, given that the research team was informed of the new survey date.

**Stage 3: Sampling of doctors (including assistant medical officers & trained nurses in the public clinics) (tertiary sampling unit)**
- All doctors (including assistant medical officers and some trained nurses in public clinics) in the sampled clinics who were on-duty on the day of survey were included.
- Locum doctors were included.
- As for doctors who are trained in clinical specialities, only family medicine specialists were included.

**Sampling of encounters**
- Record of all patient encounters seen by each health care personnel mentioned above on the survey date.

Following Figure 2.1.1 shows the study design of NMCS 2014, while Figure 2.1.2 and Figure 2.1.3 are consort diagrams which show the number of clinics sampled from each state for public and private sector respectively.
Figure 2.1.1: Study design for NMCS 2014

- **Cluster stage 1:** Random selection of clinics – primary sampling unit (PSU)
- **Cluster stage 2:** Each clinic was assigned a random survey date – secondary sampling unit (SSU)
- **Cluster stage 3:** All providers working on that date were included – tertiary sampling unit (TSU)
- **Cluster stage 4:** The encounters (all or minimum of 30) from each provider – quarternary sampling unit (QSU)
Chapter 2 : Methodology

Figure 2.1.1: Study design for NMCS 2014

Cluster stage 1: Random selection of clinics – primary sampling unit (PSU)

Cluster stage 2: Each clinic was assigned a random survey date – secondary sampling unit (SSU)

Cluster stage 3: All providers working on that date were included – tertiary sampling unit (TSU)

Cluster stage 4: The encounters (all or minimum of 30) from each provider – quarternary sampling unit (QSU)

Stratification stage 1: Sector
- Primary Care Clinics
- Public
- 14 States
- 14 States Clinics (PSU)
- Survey date (SSU)
- Providers (TSU)
- Encounters (QSU)
- Population of sampling units
- Public
- 14 States

Stratification stage 2: 14 states for each sector

Figure 2.1.2: Consort diagram – public primary care clinics 2014

Total number of public primary care clinics in 13 states and 3 federal territories
N = 911

Clinic sampling frame
N = 842

Sampled clinics
n = 139

Excluded

Closed
n = 1

Clinics without doctors
n = 68

Perlis
n = 9

Pulau Pinang
n = 5

Selangor & WP Putrajaya
n = 15

Negeri Sembilan
n = 6

Johor
n = 11

Terengganu
n = 5

Sabah & WP Labuan
n = 9

Kedah
n = 7

Perak
n = 11

WP Kuala Lumpur
n = 12

Melaka
n = 26

Pahang
n = 6

Kelantan
n = 7

Sarawak
n = 10

Closed
n = 1

Clinics without doctors
n = 68

Total number of public primary care clinics in 13 states and 3 federal territories
N = 911

Sampled clinics
n = 139

Excluded

Clinics without doctors
n = 68
Figure 2.1.3: Consort diagram – private primary care clinics 2014

Total number of private primary care clinics, N = 5,646
(CKAPS list as of 31st December 2012 matched with NHSI database)

Clinics excluded
n = 274
- Specialist clinics
- Aesthetic clinics
- In-house clinics
- Homeopathic clinics
- Operate less than 5 days per week
- Clinic closed
- Clinic shifted
- Uncontactable
- Poslaju retour
- Participated in NMCS 2012

Clinic sampling frame
N = 5,372

Sampled clinics
n = 1,002

Johor
n = 117

Kelantan
n = 45

Negeri Sembilan
n = 40

Perak
n = 73

Pulau Pinang
n = 70

Sarawak
n = 34

Pahang
n = 42

Kedah
n = 55

Melaka
n = 31

WP Kuala Lumpur
n = 120

Perlis
n = 10

Sabah & WP Labuan
n = 65

Selangor & WP Putrajaya
n = 270

Terengganu
n = 30

2.2 DATA COLLECTION AND FOLLOW-UP
The sampled clinics were each sent an invitation letter to attend a briefing in major towns in each state. Briefings for doctors in the public clinics were held on weekdays, whereas briefings for private doctors were conducted between October and December 2013 according to the convenience of the private doctors for maximum attendances. A research pack which contained the survey forms and instructions were distributed during the briefings.

To encourage further participation, representatives of clinics that did not attend the briefing were later contacted by telephone. If the doctor refused to participate, the team did not pursue further. However, if they agreed to participate the research pack was sent either by:

- courier service (Poslaju) followed by telephone call to ensure that the research kit is received. Briefing would be done over the phone to explain about the survey form
- personal visit to the clinics (within the vicinity of Klang valley), where a short private briefing would be given by the research team to the doctor/nurse in-charge

A telephone call-reminder was made to the clinic about the project and to answer any questions pertaining to the survey at two weeks and one day before the survey date. Instructions would be repeated when necessary. After the survey date, follow-up phone call(s) were made if the research pack was not returned after three weeks, and subsequently at five weeks.

Various approaches were also taken to increase the acceptance and response rates of private clinics in particular, including:

- Approaching the top management of the chain clinics/group practices.
- Obtaining a written endorsement from the Malaysian Medical Association (MMA).
- Getting support and assistance from Malaysian Medical Association (MMA) at the state level.
- Presentation of the NMCS 2012 results through general practitioners' seminar and a series of articles in MMA bulletin.
- Organising private (individual) briefings alongside Medical Practice Division's enforcement activities.

Data was collected using a self-administered questionnaire. The details of the patients managed on the date assigned to each clinic were filled by the health providers. Upon completion of data collection, participants were given certificates, which they would later use to claim for continuing professional education (CPD) points. A clinic-specific feedback, a satisfaction survey on the prescribers, and a copy of the National Medical Care Statistics 2014 report will also be sent to all participants.

2.3 RESEARCH PACK AND QUESTIONNAIRE
A pre-testing session of the questionnaire was carried out by convenience sampling of doctors from public and private clinics. The questionnaire was modified from the prior form developed in NMCS 2012 which was adapted from the Better Evaluation and Care of Health (BEACH) survey from Australia. A total of 30 encounters were recorded, and comments from the doctors based on the pre-testing were taken into consideration to further improvise the form. The NMCS 2014 form was modified based on these feedbacks and the finalised form is enclosed in Appendix 3.
2.2 DATA COLLECTION AND FOLLOW-UP

The sampled clinics were each sent an invitation letter to attend a briefing in major towns in each state. Briefings for doctors in the public clinics were held on weekdays, whereas briefings for private doctors were conducted between October and December 2013 according to the convenience of the private doctors for maximum attendances. A research pack which contained the survey forms and instructions were distributed during the briefings.

To encourage further participation, representatives of clinics that did not attend the briefing were later contacted by telephone. If the doctor refused to participate, the team did not pursue further. However, if they agreed to participate the research pack was sent either by:

- courier service (Poslaju) followed by telephone call to ensure that the research kit is received. Briefing would be done over the phone to explain about the survey form
- personal visit to the clinics (within the vicinity of Klang valley), where a short private briefing would be given by the research team to the doctor/nurse in-charge

A telephone call-reminder was made to the clinic about the project and to answer any questions pertaining to the survey at two weeks and one day before the survey date. Instructions would be repeated when necessary. After the survey date, follow-up phone call(s) were made if the research pack was not returned after three weeks, and subsequently at five weeks.

Various approaches were also taken to increase the acceptance and response rates of private clinics in particular, including:

a) Approaching the top management of the chain clinics/group practices.
b) Obtaining a written endorsement from the Malaysian Medical Association (MMA).
c) Getting support and assistance from Malaysian Medical Association (MMA) at the state level.
d) Presentation of the NMCS 2012 results through general practitioners’ seminar and a series of articles in MMA bulletin.
e) Organising private (individual) briefings alongside Medical Practice Division’s enforcement activities.

Data was collected using a self-administered questionnaire. The details of the patients managed on the date assigned to each clinic were filled by the health providers. Upon completion of data collection, participants were given certificates, which they would later use to claim for continuing professional education (CPD) points. A clinic-specific feedback, a satisfaction survey on the prescribers, and a copy of the National Medical Care Statistics 2014 report will also be sent to all participants.

2.3 RESEARCH PACK AND QUESTIONNAIRE

A pre-testing session of the questionnaire was carried out by convenience sampling of doctors from public and private clinics. The questionnaire was modified from the prior form developed in NMCS 2012 which was adapted from the Better Bettering the Evaluation and Care of Health (BEACH) survey from Australia. A total of 30 encounters were recorded, and comments from the doctors based on the pre-testing were taken into consideration to further improvise the form. The NMCS 2014 form was modified based on these feedbacks and the finalised form is enclosed in Appendix 3.
Each research pack contained:

- **Survey pads**
  - 40 forms
  - One set of instructions
  - One case study
  - One example of a completed form

- **NMCS 2014 summary information**
  - Objectives of the NMCS 2014
  - Brief description of project and project team
  - Individual survey date of the clinics

- **Public notice**
  - Notice to be displayed in the participating clinic to inform patients that the clinic is currently undertaking the NMCS survey

- **ICPC-2-code list**
  - ICPC-2-Code list

Also included in the research pack:

- **Call letter**
  - Letter signed by the Director of the State Health Departments to inform the participating clinics of the survey

- **Prepaid envelope**
  - One envelope for every two survey pads

### 2.4 DATA MANAGEMENT

#### Data entry

Prior to the start of data entry, all data entry personnel were given reference materials containing a description of the study, examples of the questionnaire, classification and coding systems, data entry rules and regulations. This was followed by two sessions of data entry training of at least 2 hours each session. Data is then transferred from paper to an electronic format through a data entry web application by trained data entry personnel.

**Session 1: Demonstration and practical session**

- Slide presentation on data entry module
- Live demonstration of data entry module
- Live demonstration of coding systems
- Discussion on data entry and coding systems
- Practical session – practice data entry and coding of 20 test questionnaires per data entry personnel
Chapter 2: Methodology

Each research pack contained:

- Survey pads
  - 40 forms
  - One set of instructions
  - One case study
  - One example of a completed form

- NMCS 2014 summary information
  - Objectives of the NMCS 2014
  - Brief description of project and project team
  - Individual survey date of the clinics

- Public notice
  - Notice to be displayed in the participating clinic to inform patients that the clinic is currently undertaking the NMCS survey

- ICPC-2-code list

Also included in the research pack:

- Call letter
  - Letter signed by the Director of the State Health Departments to inform the participating clinics of the survey

- Prepaid envelope
  - One envelope for every two survey pads

2.4 DATA MANAGEMENT

Data entry

Prior to the start of data entry, all data entry personnel were given reference materials containing a description of the study, examples of the questionnaire, classification and coding systems, data entry rules and regulations. This was followed by two sessions of data entry training of at least 2 hours each session. Data is then transferred from paper to an electronic format through a data entry web application by trained data entry personnel.

Session 1: Demonstration and practical session

- Slide presentation on data entry module
- Live demonstration of data entry module
- Live demonstration of coding systems
- Discussion on data entry and coding systems
- Practical session – practice data entry and coding of 20 test questionnaires per data entry personnel

Session 2: Question and answer

- Feedback was provided to data entry personnel on data entry and coding issues from the 20 test questionnaires

Standardisations to the data entry rules and coding systems were also periodically updated and conveyed to all data entry personnel.

Data quality assurance

The data entry application was loaded with previous coding history from NMCS 2012 and also current coding entry to ease the coding process and to ensure consistency of coding.

Software based quality assurance measures were also built into the data entry applications either as a quality measure or to facilitate the data entry process. For example warnings prompt when there was a duplication of identification card number being entered, warnings prompt of missing mandatory fields, auto-generation of date of birth and age through identification card number when available etc.

Validity checks were put in place during data entry to minimise entry of illogical data and warnings would pop-up if extreme values were entered to prompt the data entry personnel to re-check the data. These include validation on the date of birth entered, gender counter check via identification card number when available, unable to enter the same diagnosis within the same encounter etc.

In addition to the aforementioned measures, double data entry was also incorporated as part of the quality assurance of the data. This form of quality check has been recommended and known to correct data entry errors from the original entry.5

Double data entry was done for more than 10% of the total entries (2,894 out of 27,808 forms) in six batches, where Batch 1 was completed in June 2014 and Batch 6 in October 2014. Questionnaires that were to be entered a second time were identified by random selection of clinics. The data entry personnel were blinded to the assignment of clinics for double data entry.

For each batch of double data entry, all discrepancies between the first and second set of records were verified by checking either with the original forms or the coding definitions. Errors were defined as deviations of either the first or second entry from the original questionnaire by alphanumeric characters or assigning the wrong code for a variable. However those errors that were due illegible handwriting were not regarded as an error. A correct third record was then updated into the database.

The percentage of data entry error for each available variable was then calculated by obtaining the proportion of errors per total cases within the variable. The variables with the highest rates of data entry error were then compared.
Table 2.4.1: Data entry error rate for NMCS 2014

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coded variables</th>
<th>Data entry error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ICPC-2+ code</td>
<td>Batch 1</td>
</tr>
<tr>
<td></td>
<td>ATC code</td>
<td>3.5</td>
</tr>
<tr>
<td>Non-coded variables</td>
<td>Nationality</td>
<td>16.3</td>
</tr>
<tr>
<td></td>
<td>Procedures/other treatments/ counselling</td>
<td>8.9</td>
</tr>
<tr>
<td></td>
<td>Diagnosis not specified for which medical certificated was issued</td>
<td>6.2</td>
</tr>
</tbody>
</table>

The three variables for the non-coding section were the variables with the highest data entry error rate for batch 1. There was marked improvement in error rate for these variables from batches 1 to 5. Increase of data entry error rate for the coded variables can be attributed to recruitment of new data entry personnel, resulting in more variations in coding. While many of the errors were random errors but coding errors were largely occurring in a systematic manner; where a data entry personnel with a misconception of the correct codes for certain diseases/medications, makes a consistent error throughout all forms entered.

There does not appear to have a general consensus of acceptable data entry error rate worldwide. Previous study shown that error rates detected by double-entry method for clinical databases ranged from 2.3 to 5.2% for demographic data while for treatment data, it ranged from 10.0 to 26.9%. Similarly, Fontaine et. al reported an overall rate of 7.3% for data entry strategies used in clinical trial.

Double entry has been recognised as the gold standard in transferring of data into an electronic database but it substantially increases the amount of time and costs of data entry. Costs of resources have been reported to be increased by up to 2.5 times with double data entry compared to single entry. Also, additional software solutions and manual checking mechanisms are required when performing checks on discrepancies and putting in corrections.

An alternative recommendation is a trade-off between acceptable data accuracy and cost-effectiveness using single data entry with concurrent quality control measures, exploratory data analysis and post-entry logic checks. It is also recognised that double entry detects errors where exploratory analysis misses while on the other hand not all discrepancies found by exploratory data analysis is identified by double entry. Hence, suggests that double data entry alone may not necessarily be sufficient as a sole data quality checking method.

All the errors which were detected (coded and non-coded) were corrected by referring to the original forms and by discussion among the investigators and the Research Evaluation Committee. Further logic checks and exploratory analyses were also conducted during data cleaning to question the plausibility and ensure the validity of the data. A protocol with validation rules for cleaning as well as data inconsistency rules was compiled for the purpose of data cleaning.
Classification of data (data coding)

International Classification of Primary Care (ICPC)

The International Classification of Primary Care Second Edition (ICPC-2) was used to classify the following data elements:

- Reasons for encounter
- Diagnoses
- Investigations
- Procedures
- Advice/counselling

The ICPC-2 is accepted by the World Health Organization (WHO) as a member of the WHO Family of International Classifications. It was published in 1987 by the World Organisation of Family Doctors (WONCA) and used in more than 45 countries as the standard for data classification in primary care. The ICPC-2 has a bi-axial structure, with 17 chapters based on body systems (Table 2.4.2) and seven components (Table 2.4.3) with rubrics bearing a letter and two-digit numeric code.

The data were entered and coded using ICPC-2 PLUS, an extended clinical terminology classified according to ICPC-2. ICPC-2 PLUS coding system contains extended terms commonly used in general practice that are more specific, and helps to ensure accurate classification to ICPC-2 during data entry. ICPC-2 PLUS was developed in 1995, and is maintained and regularly updated by the Family Medicine Research Centre (FMRC) of the University of Sydney. Also known as BEACH coding system, ICPC-2 PLUS is primarily used in Australia especially for the national study of general practice activity, the BEACH program.

Table 2.4.2: ICPC-2 chapters

<table>
<thead>
<tr>
<th>Code</th>
<th>ICPC-2 chapter</th>
<th>Code</th>
<th>ICPC-2 chapter</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>General</td>
<td>B</td>
<td>Blood, immune system</td>
</tr>
<tr>
<td>D</td>
<td>Digestive</td>
<td>F</td>
<td>Eye</td>
</tr>
<tr>
<td>H</td>
<td>Ear</td>
<td>K</td>
<td>Circulatory</td>
</tr>
<tr>
<td>L</td>
<td>Musculoskeletal</td>
<td>N</td>
<td>Neurological</td>
</tr>
<tr>
<td>P</td>
<td>Psychological</td>
<td>R</td>
<td>Respiratory</td>
</tr>
<tr>
<td>S</td>
<td>Skin</td>
<td>T</td>
<td>Endocrine, nutritional &amp; metabolic</td>
</tr>
<tr>
<td>U</td>
<td>Urological</td>
<td>W</td>
<td>Women's health, pregnancy, family planning</td>
</tr>
<tr>
<td>X</td>
<td>Female genital</td>
<td>Y</td>
<td>Male genital</td>
</tr>
<tr>
<td>Z</td>
<td>Social problems</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2.4.3: ICPC-2 components

<table>
<thead>
<tr>
<th>ICPC-2 components</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Complaints and symptoms</td>
<td>01–29</td>
</tr>
<tr>
<td>2. Diagnostics, screening and preventive</td>
<td>30–49</td>
</tr>
<tr>
<td>3. Medication, treatment, procedures</td>
<td>50–59</td>
</tr>
<tr>
<td>4. Test results</td>
<td>60–61</td>
</tr>
<tr>
<td>5. Administrative</td>
<td>62</td>
</tr>
<tr>
<td>6. Referrals</td>
<td>63–69</td>
</tr>
<tr>
<td>7. Diagnoses, diseases</td>
<td>70–99</td>
</tr>
<tr>
<td>– infectious</td>
<td></td>
</tr>
<tr>
<td>– neoplastic</td>
<td></td>
</tr>
<tr>
<td>– injuries</td>
<td></td>
</tr>
<tr>
<td>– congenital anomalies</td>
<td></td>
</tr>
<tr>
<td>– others</td>
<td></td>
</tr>
</tbody>
</table>

The National Clinical Research Centre has been granted a free research licence from WONCA for the usage of ICPC-2 codes in the NHSI project which is valid from February 2011 till end of 2014 whereas the ICPC-2 PLUS was obtained under a free licence from the University of Sydney.

Results were reported at the ICPC-2 classification level. Some of the diagnoses were grouped together by combining several ICPC-2 or ICPC-2 PLUS codes (Appendix 4). Classification of pathology and imaging test according to ICPC-2 can be very broad (e.g. HbA1c test is classified under T34 - Blood test endo/metabolic). Hence, results for Chapter 10 were presented as ICPC-2 PLUS.

**Anatomical Therapeutic Chemical (ATC) classification**

Medications were coded and classified using the Anatomical Therapeutic Chemical (ATC) classification system. ATC has been recommended by the WHO and used in many countries including Malaysia, as a global standard for classifying medications for drug utilisation research, evaluating trend of drug consumption and for international comparisons. Medications are classified into groups at five different levels, with the following example:

- Level 1: C - Cardiovascular system
- Level 2: C10 - Serum lipid reducing agents
- Level 3: C10A - Cholesterol of triglyceride reducers
- Level 4: C10AA - HMG CoA reductase inhibitors
- Level 5: C10AA01 – Simvastatin

The ATC licence was purchased from the WHO Collaborating Centre for Drug Statistics Methodology. Medications were entered as free text in generic (non-proprietary) or brand name, and coded by trained data entry personnel according to the Guidelines for ATC Classification and DDD assignment 2012. In certain cases, the doctors might not specify the medications down to the generic level hence it could only be coded to ATC level 3 or 4.
2.5 DATA ANALYSIS

Weighting

The data presented in this report were weighted to adjust for over and under representativeness of any strata in the sample as well as to account for non-respondents. Table 2.5.1 shows the 28 weighting strata that were defined for the study population, by state/region and sector. The components incorporated in the estimation of total weights are described below.

Table 2.5.1: Strata according to state/region and sector

<table>
<thead>
<tr>
<th>State/federal territory</th>
<th>Sector</th>
<th>Stratum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Johor</td>
<td>Public</td>
<td>J1</td>
</tr>
<tr>
<td></td>
<td>Private</td>
<td>J2</td>
</tr>
<tr>
<td>Kedah</td>
<td>Public</td>
<td>K1</td>
</tr>
<tr>
<td></td>
<td>Private</td>
<td>K2</td>
</tr>
<tr>
<td>Kelantan</td>
<td>Public</td>
<td>D1</td>
</tr>
<tr>
<td></td>
<td>Private</td>
<td>D2</td>
</tr>
<tr>
<td>Melaka</td>
<td>Public</td>
<td>M1</td>
</tr>
<tr>
<td></td>
<td>Private</td>
<td>M2</td>
</tr>
<tr>
<td>Negeri Sembilan</td>
<td>Public</td>
<td>N1</td>
</tr>
<tr>
<td></td>
<td>Private</td>
<td>N2</td>
</tr>
<tr>
<td>Pahang</td>
<td>Public</td>
<td>C1</td>
</tr>
<tr>
<td></td>
<td>Private</td>
<td>C2</td>
</tr>
<tr>
<td>Perak</td>
<td>Public</td>
<td>A1</td>
</tr>
<tr>
<td></td>
<td>Private</td>
<td>A2</td>
</tr>
<tr>
<td>Perlis</td>
<td>Public</td>
<td>R1</td>
</tr>
<tr>
<td></td>
<td>Private</td>
<td>R2</td>
</tr>
<tr>
<td>Pulau Pinang</td>
<td>Public</td>
<td>P1</td>
</tr>
<tr>
<td></td>
<td>Private</td>
<td>P2</td>
</tr>
<tr>
<td>Sabah &amp; WP Labuan</td>
<td>Public</td>
<td>SB1</td>
</tr>
<tr>
<td></td>
<td>Private</td>
<td>SB2</td>
</tr>
<tr>
<td>Sarawak</td>
<td>Public</td>
<td>SW1</td>
</tr>
<tr>
<td></td>
<td>Private</td>
<td>SW2</td>
</tr>
<tr>
<td>Selangor &amp; WP Putrajaya</td>
<td>Public</td>
<td>B1</td>
</tr>
<tr>
<td></td>
<td>Private</td>
<td>B2</td>
</tr>
<tr>
<td>WP Kuala Lumpur</td>
<td>Public</td>
<td>W1</td>
</tr>
<tr>
<td></td>
<td>Private</td>
<td>W2</td>
</tr>
</tbody>
</table>
**Sampling weight**

Sampling weight is the inverse of the probability of selecting a unit.\(^{13}\) The sampling weight of each stratum calculated as follow:\(^{14}\)

\[
SW_j = \frac{M_j}{m_{j,\text{res}} + m_{j,\text{non}} + m_{j,\text{exc}}}
\]

where \(M_j\) is the total number of primary care clinics that can be sampled in the \(j^{th}\) strata (population), \(m_{j,\text{res}}\) is the number of primary care clinics responded for strata \(j\), \(m_{j,\text{non}}\) is the number of primary care clinics who did not respond in the \(j^{th}\) strata, and \(m_{j,\text{exc}}\) is the number of clinics excluded after being sampled for strata \(j\).

**Activity weight**

The activity weight for each clinic was calculated to account for the different level of activities of each clinic. It was calculated as follows:

\[
AW_{jk} = \frac{N_{jk}}{n_{jk}}
\]

where \(N_{jk}\) is the expected patients' visits per day of the \(k^{th}\) clinic in the \(j^{th}\) strata while \(n_{jk}\) is the number of encounters we received from the \(k^{th}\) clinic in the \(j^{th}\) strata.

**Adjustment for non-response**

To account for less than 100% response rate, adjustment for the non-response is required.\(^{12}\) The non-response adjustment weight was calculated as follows:

\[
A_j = \frac{m_{j,\text{res}} + m_{j,\text{non}}}{m_{j,\text{res}}}
\]

where \(m_{j,\text{res}}\) is the number of primary care clinics responded for strata \(j\) and \(m_{j,\text{non}}\) is the number of primary care clinics who did not respond in the \(j^{th}\) strata.

**Total weight**

The final weight for each stratum was calculated as the multiplication of the sampling weight, activity weight and adjustment for non-response.

\[
FW_{jk} = SW_j \times AW_{jk} \times A_j
\]

The weighted estimates were generated using the survey package in R.
Sampling weight

Sampling weight is the inverse of the probability of selecting a unit. The sampling weight of each stratum calculated as follow:

\[ \text{Sampling weight of each stratum} = \frac{M_j}{m_{j,\text{res}}} + \frac{m_{j,\text{non}}}{m_{j,\text{exc}}} \]

where \( M_j \) is the total number of primary care clinics that can be sampled in the \( j \)th strata (population), \( m_{j,\text{res}} \) is the number of primary care clinics responded for strata \( j \), \( m_{j,\text{non}} \) is the number of primary care clinics who did not respond in the \( j \)th strata, and \( m_{j,\text{exc}} \) is the number of clinics excluded after being sampled for strata \( j \).

Activity weight

The activity weight for each clinic was calculated to account for the different level of activities of each clinic. It was calculated as follows:

\[ \text{Activity weight of each clinic} = \frac{N_{jk}}{n_{jk}} \]

where \( N_{jk} \) is the expected patients' visits per day of the \( k \)th clinic in the \( j \)th strata while \( n_{jk} \) is the number of encounters we received from the \( k \)th clinic in the \( j \)th strata.

Adjustment for non-response

To account for less than 100% response rate, adjustment for the non-response is required. The non-response adjustment weight was calculated as follows:

\[ \text{Non-response adjustment weight} = \frac{m_{j,\text{res}}}{m_{j,\text{res}} + m_{j,\text{non}}} \]

Total weight

The final weight for each stratum was calculated as the multiplication of the sampling weight, activity weight and adjustment for non-response.

\[ \text{Total weight} = \text{Sampling weight} \times \text{Activity weight} \times \text{Non-response adjustment weight} \]

The weighted estimates were generated using the survey package in R.

Statistical analysis

Analysis was done in R with an R package called "survey: analysis of complex survey samples". Results are presented as number of unweighted counts, weighted counts, proportions and rate per 100 encounters along with 95% confidence interval (CI). Rate per 100 diagnoses are reported for management that can occur at more than once per diagnosis.

2.6 ETHICS APPROVAL

The study was approved by the Medical Research and Ethics Committee (MREC) (Approval Number: NMRR-09-842-4718). As per previous study, a public notice was placed at each participating clinic to inform patients that their prescription data would be collected for research purposes. Patients had the right to decline to participate at any point of time throughout the study period.

2.7 LIMITATIONS

1. The survey is self-administered and therefore precision of data depends largely on the completeness of recording by respondents, hence may not accurately reflect true practice.
2. The survey is encounter-based and reflects the morbidity pattern observed in the primary care setting rather than the prevalence of disease in the community.
3. The morbidity patterns reflect only those morbidities managed during the recorded encounters. There may be co-morbidity in the same patient which was not expected to be managed during the encounter and hence was not recorded.
4. This is a cross-sectional study. Therefore, no conclusions may be generated on the outcomes of management of acute and chronic diseases in the primary care setting. Prescriptions, procedures, imaging and referrals reported were those provided at the present point of encounter and did not necessarily indicate that the patient has not already received them in a previous encounter.
5. Maternal child health encounters in public clinics were mostly attended by trained nurses. NMCS 2014 might miss those cases as not all the trained nurses were involved in the study.
6. The sampling of public clinics can be improved by incorporating the classification of the type of clinics, which is based on the workload of the clinic.
7. Verification of data received via audit process was not done. All data received were presumed to be accurate and precise.
8. Benchmarking the sample against population data cannot be performed as there is no readily available primary care population data, be it the providers or the patients.
9. Non-respondent details were not recorded; hence non-response analysis to compare the sample and the non-respondent cannot be performed.
REFERENCES