

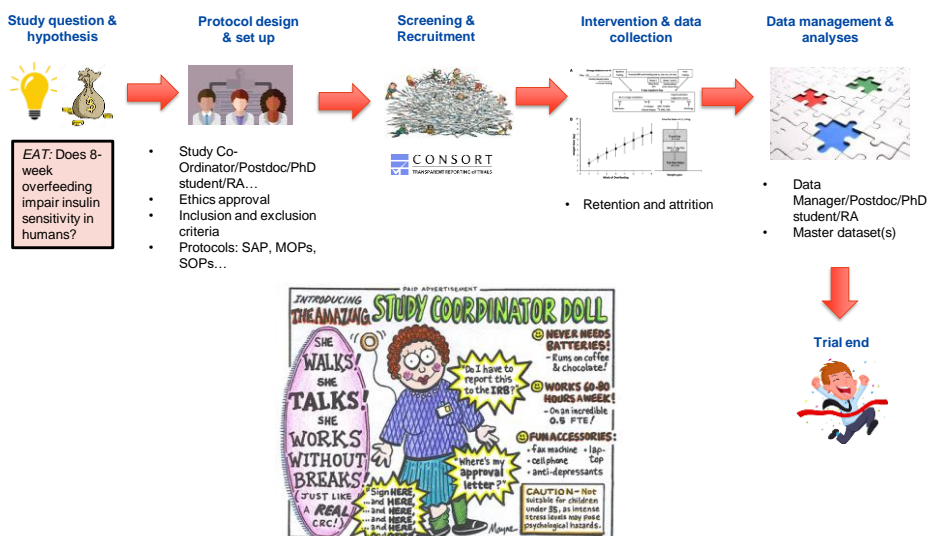
Harnessing digital technologies for clinical trials research

Charmaine Tam, PhD
Senior Research Fellow &
Analytics Translator

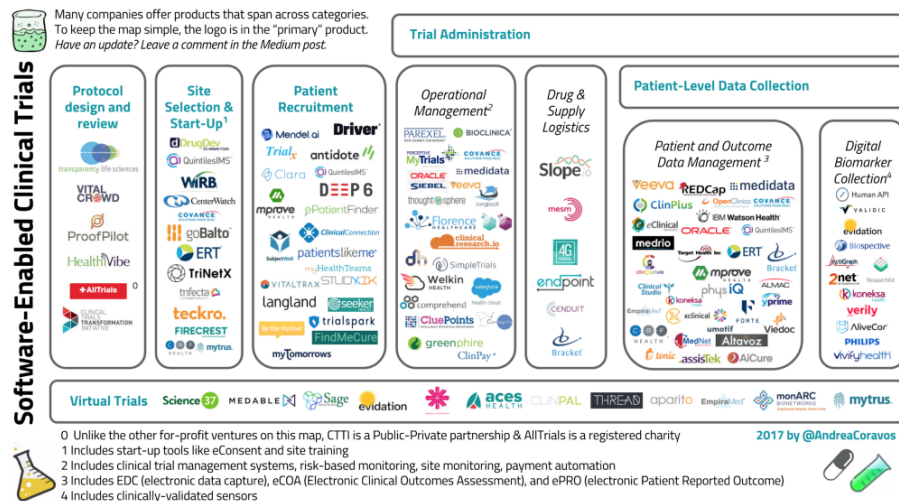
Northern Clinical School, Faculty of Medicine and Health
Centre for Translational Data Science



Behind the scenes of a clinical trial, circa 2012



Incorporating digital technology into clinical trial workflow

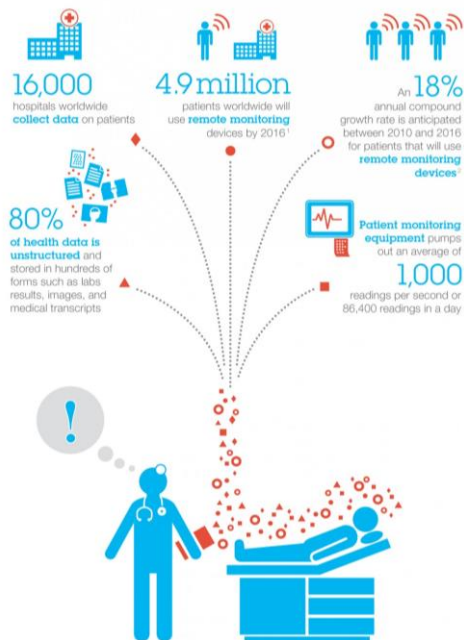


Increasing patient recruitment in ANZ

<https://digital.hbs.edu/innovation-disruption/software-enabled-clinical-trials/>
 Published 30th September 2019

The University of Sydney

Paradigm shift is enabled by ever increasing sources of digital health data



Opportunities in clinical trials

- Identification and recruitment of eligible patients using computable phenotypes
- Improved retention and attrition of patients
- Reduced costs of data collection
- Scalability and reproducibility
- Improved representation from under-represented populations
- Follow up of patients beyond clinical trials

<https://www.ibmdatahub.com/infographic/big-data-healthcare-tapping-new-insight-save-lives>

Page 4

Comparing data from clinical trials with other sources of data

	Clinical trial	Administrative Data Sets (eg. Claims, Admitted, Non-Admitted Patient collections, etc)	eMR
Original purpose of data collection	Research Primary data collection	Activity-based funding Secondary data collection	Clinical Care Secondary data collection
Common study design	Cross-sectional, prospective cohort	Retrospective, cross-sectional	Cross-sectional, retrospective or prospective cohort
Cost	\$\$\$\$ primarily government funded	\$\$ primarily government funded	\$ data collection is funded by health care systems; research can be funded by a variety of sources or may not require funding at all
Access	Researchers with ethics and governance approval	From data warehouse by analysts/data managers in hospitals/industry, researchers with ethics and governance approval	Researchers with ethics and governance approval
Time frame	Follow-up is restricted by funding; must wait for health outcomes to occur for longitudinal studies	Retrospective with datasets updated daily	Retrospective restricted by date of eMR implementation; additional data collected at low cost
Study population and follow-up	Based on recruitment, follow-up is scheduled	Curated and aggregated data from eMR systems across a specified population (eg. NSW)	Based on patient use of the specific health system and information system
Data collection and storage	Established protocols with robust, well-documented approaches to data collection	Established protocols with robust, well-documented approaches to data collection (eg. from eMR)	Little to no established protocols and no data dictionary
Conditions captured	Any outcomes as specified at the beginning of the study	Captures coded diagnoses and procedures and selected outcomes (eg. length of stay, in-hospital mortality)	Only outcomes requiring care by a clinician, missingness of data due to non-recording
Summary of pros and cons	High quality data in a pre-specified study population, no-real time, expensive to maintain and extend	State-wide coverage, updated daily and used for reporting, doesn't capture clinically meaningful information	Potential to be used in real-time applications, no established protocols for data collection or quality, big data-volume, velocity, variety, veracity, value, scalable and reproducible and cost-effective

+ data collected from devices (eg. wearables, apps)

The University of Sydney

Page 5

SPEED-EXTRACT (STEMI Patient EElectronic Data EXTRACTION)

Proof-of-concept project funded by Ministry of Health July 2018-December 2019

Primary Aim

To demonstrate the feasibility of accurately identifying (>90%) patients with ST Elevation Myocardial Infarction (STEMI) from existing suspected Acute Coronary Syndrome data, that reside in electronic medical record systems (EMR) from one quaternary and two feeder hospitals from Northern Sydney LHD within the Sydney Health Partners collaborative.

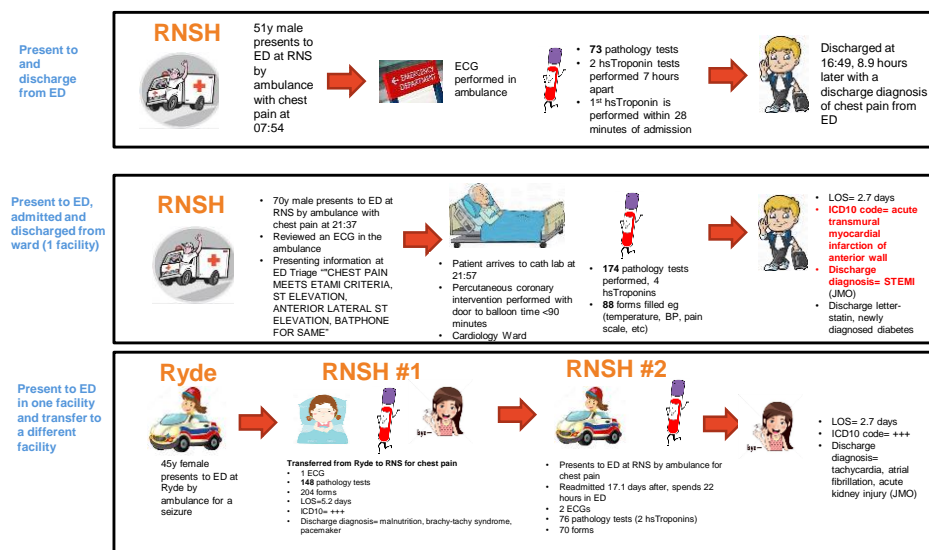
Specific aims

1. To develop a process and method for standardised extraction of eMR data to identify patients with a discharge diagnosis of (STEMI)
2. To determine the extent to which extracted data elements can be used to describe clinical quality measures
3. To share this identified cohort and proposed quality measures with practicing clinicians to ensure 'face validity' of the extracted data

The University of Sydney

Page 6

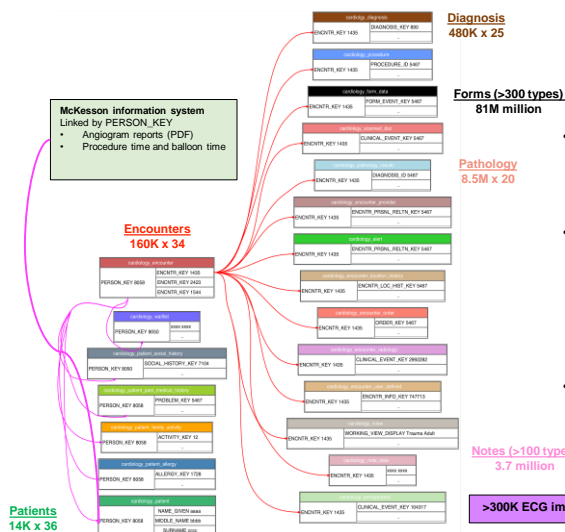
Mapping data extraction to reconstruct patient journeys for patients with chest pain



The University of Sydney

Page 7

Overview of data in eMR systems



Data extracted from the eMR is structured in transactional data format. ie. Every transaction/entry that occurs in the hospital is recorded as one line item

To reconstruct a patient's journey through hospital, tables need to be linked by:

- Encounter_ID**= unique identifier for a patient's interaction with the hospital eg. emergency, inpatient, outpatient, community) AND/OR
- Person_ID**= unique patient identifier

For every event of interest, we are examining the most appropriate time-stamp to describe an event

The University of Sydney

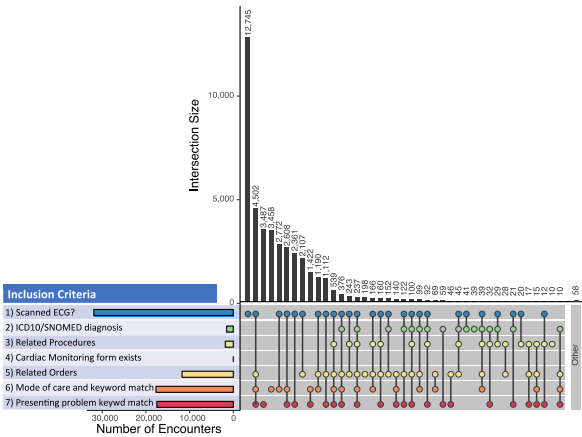
Page 8

Identifying all presentations of suspected acute coronary syndrome

- 3 months Test Data Extract: April-June 2017
- Bulk Data Extract: January 2013- June 2018
- Historical and future encounters are extracted

Cardiac keywords and symptoms
Chest pain, chest tightness, shortness of breath, dyspnoea, weakness, nausea, vomiting, palpitations, syncope, presyncope, unwell, cardiac arrest, indigestion, sweaty, diaphoresis, dizziness, light-headedness, fatigue, clamminess, pale, ashen, loss of consciousness, SALAMI, ETAMI, STEMI, NSTEMI, out of hospital cardiac arrest, ventricular tachycardia, ventricular fibrillation, failed thrombolysis, cath, cath lab, coronary bypass graft, ami, stent, angiogram, angio, epigastric pain, arm heaviness, chest heaviness

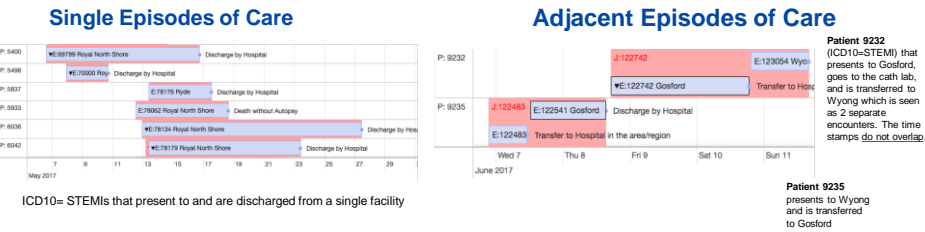
Included abbreviations, misspellings and additional keywords



UpSet plot showing the numbers of encounters meeting individual (left hand side) and multiple inclusion criteria (right hand side)

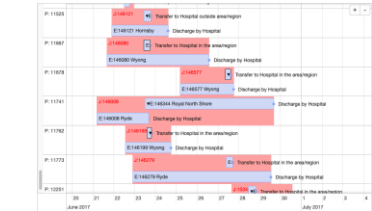
Diversity of episodes of care

Data in the EMR is captured as encounters. To enable patient-centric analyses, this data was converted into episodes of care. This is especially important for transferred patients where data is captured as ≥2 encounters



ICD10= STEMI= STEMI that present to and are discharged from a single facility

Overlapping Episodes of Care



The time stamps do overlap between encounters

Each row depicts a single patient. The pink shading represents the new journeyID that's been created. Blue boxes represent encounters. ♥ is a proxy for cath lab (admission to or procedure performed in cath lab)

Validation study of ICD10 coded STEMI

- **Rationale:** ICD10 codes can be used to identify STEMI but are not entirely reliable and are only available after the episode of care
- Designed and built a user interface where cardiologists can easily sight all relevant aspects of *N* patient records (one at a time) and select a diagnosis. Data includes:
 - ECGs
 - First medical note
 - Blood tests (incl. hsTroponin)
 - Angiogram report
 - Discharge letter
- **Population for Validation Study**
 - The starting population is 1144 episodes of care from admitted patients in NSCCLHD with hsTroponin changes*
 - Of these we will select 750 unique episodes of care for validation which will include cases with and without ICD10=STEMI

Outcome
Labelled dataset that can be used to train algorithm(s) to identify "real" STEMI

The University of Sydney

* a) Δ 30% between initial and subsequent hsTroponin measurements AND b) at least one hsTroponin measured during the encounter is >99th percentile for normal reference population OR If hsTroponin > 1000ng/L

Page 11



Acute Coronary Syndromes Clinical Care Standard

- 1 A patient presenting with acute chest pain or other symptoms suggestive of an acute coronary syndrome receives care guided by a documented chest pain assessment pathway.
- 2 A patient with acute chest pain or other symptoms suggestive of an acute coronary syndrome receives a 12-lead electrocardiogram (ECG) and the results are analysed by a clinician experienced in interpreting an ECG within 10 minutes of the first emergency clinical contact.
- 3 A patient with an acute ST-segment-elevation myocardial infarction (STEMI), for whom emergency reperfusion is clinically appropriate, is offered timely percutaneous coronary intervention (PCI) or fibrinolysis in accordance with the time frames recommended in the current National Heart Foundation of Australia/Cardiac Society of Australia and New Zealand Guidelines for the Management of Acute Coronary Syndromes.
In general, primary PCI is recommended if the time from first medical contact to balloon inflation is anticipated to be less than 90 minutes, otherwise the patient is offered fibrinolysis.
- 4 A patient with a non-ST-segment-elevation acute coronary syndrome (NSTEMACS) is managed based on a documented, evidence-based assessment of their risk of an adverse event.
- 5 The role of coronary angiography, with a view to timely and appropriate coronary revascularisation, is discussed with a patient with a non-ST-segment-elevation acute coronary syndrome (NSTEMACS) who is assessed to be at intermediate or high risk of an adverse cardiac event.
- 6 Before a patient with an acute coronary syndrome leaves the hospital, they are involved in the development of an individualised care plan. This plan identifies the lifestyle modifications and medicines needed to manage their risk factors, addresses their psychosocial needs and includes a referral to an appropriate cardiac rehabilitation or another secondary prevention program. This plan is provided to the patient and their general practitioner or ongoing clinical provider within 48 hours of discharge.

Acute Coronary Syndromes Guidelines Writing Group. Guidelines for the management of acute coronary syndromes 2008. Medical Journal of Australia, 2008; 194(5): 52-53.

Indicators that are potentially measurable using EMR data

Quality statement 2 – Early Assessment

- **Indicator 2b:** ECG performed within 10 minutes of arrival of ambulance
- **Indicator 2c:** ECG performed and interpreted within 10 minutes of arrival to ED

✓

Quality statement 3 – Timely Reperfusion

- **Indicator 3a:** STEMI patients receiving fibrinolysis or PCI
- **Indicator 3b:** STEMI patients receiving fibrinolysis within 30 minutes of hospital arrival
- **Indicator 3c:** PCI patients with STEMI with door-to-device within 90 minutes

✓

Quality statement 6 – Individualised Care Plan

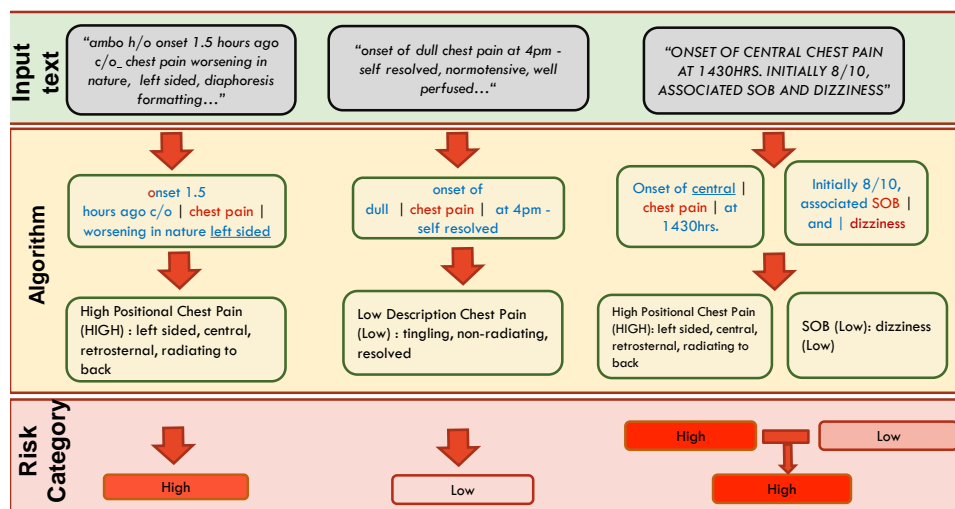
- **Indicator 6b:** Patients discharged on aspirin or dual antiplatelet therapy

✓

The University of Sydney

Page 12

Step 2: Text Mining Contextual Analysis Examples



The University of Sydney

Page 17

Symptom/Keyword Grouping and Risk Category

Symptom/Keyword Group	Risk Category
high_positional_chest_pain	High
chest_tightness/heaviness	High
check_coronary_artery_symptom	High
high_description_chest_pain	High
check_squeezing_constricted_banding	High
check_ecg_changes	High
check_etami	High
check_axilla_pain	High
check_retros_heaviness	High
intermediate_description_chest_pain	Intermediate
check_clopidogrel	Intermediate
check_pain_on_exertion	Intermediate
shortness_of_breath	Low
dizzy	Low
vomiting	Low
nausea	Low
pale	Low
palpitation	Low
no chest pain	No

- Developed text mining algorithm in the entire cohort of patients with suspected acute coronary syndrome
- 36 different symptom/keyword searches which accommodates for abbreviations and misspellings
- The encounter is assigned the highest risk category associated with the symptom/keyword group identified in the text

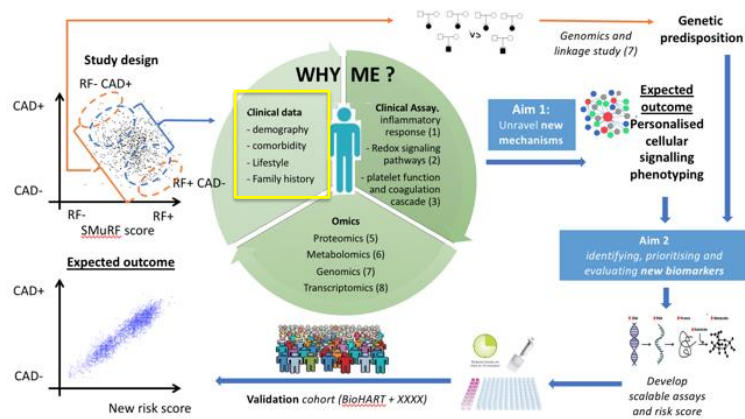
Our text mining tool called "CLACK" can take any input text source and examine strings of interest eg. risk factors, evidence of thrombolysis, past medical history, etc

The University of Sydney

Page 18

Proof-of-concept project to automate extraction of clinical data from eMR to enhance discovery of new biomarkers of coronary artery disease

Aim: To test the feasibility of using eMR data from SPEED-EXTRACT to populate BioHEART, an ongoing clinical trial to identify new biomarkers and new mechanisms of coronary artery disease and myocardial infarction in patients presenting with STEMI



Mapping data variables between clinical trials and eMR data sources

Example of risk factors

BioHEART Clinical Trial	SPEED-EXTRACT eMR
Hypertension?	1. Prior medical history of hypertension recorded in ED notes (any occurrence of htn, hypertension, excluding negations) 2. New record of hypertension recorded in progress notes (as above) 3. Existing medications for hypertension (1077 types) 4. Mean of 10 highest BP measurements over the episode of care
Current Smoker?	1. Record of ex- or current smoker recorded in structured field 2. Record of ex- or current smoker in ED notes
Previous Smoker?	1. Record of ex- or current smoker recorded in structured field 2. Record of ex- or current smoker in ED notes
Pack Years (20/day)	Record of "pack years" recorded in ED notes
Dyslipidaemia?	1. Prior medical history of hyperlipidemia recorded in ED notes (any occurrence of hdl, hypercholesterolaemia, excluding negations) 2. New record of hyperlipidemia in progress notes (as above) 3. Prior prescription for statin or lipid-lowering medications (518 types) 4. abnormal cholesterol test results (pathology)
Diabetes?	1. A HbA1c test result greater than 6.5 percent 2. Prior medical history of diabetes recorded in ED notes (any occurrence of diabetes, diabetic, T2DM, IDDM, NIDDM, excluding negations). 3. New occurrence of diabetes recorded in progress notes (as above) 4. Any record of medications for insulin or antidiabetics (83 types)

+ presenting symptoms, family history, medical history, complications....

The ABC of embedded trials (Prof Neena Modi)

Accurate data

- Clear oversight and transparency across the data pipeline from data extraction from information systems, processing and analysis
- Ensuring data quality by following the data journey from ED through to medical coding

Benefits and buy-in

- Fortnightly meetings with stakeholders spanning Ministry of Health, eHealth, cardiologists, professors in nursing and digital health, software engineers, data analysts and data scientists

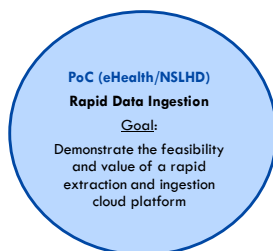
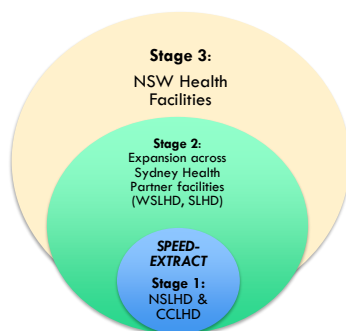
Confidence and trust

- Embedded in the health system closest to the data source
- Communication and iteration of results to check face validity

The University of Sydney

Page 21

Work in progress



Accord for the Sharing of Health Data across Sydney Health Partners



PoC to test whether this dataset can be updated in close to real-time

Ongoing work

- Data pipeline
- Assessment and management of diabetes in patients with acute MI
- Health care utilisation in patients presenting with low risk of cardiac related chest pain
- Risk factors in patients with STEMI
- BNP levels in heart failure...

The University of Sydney

Page 22

It takes a herd!!



The University of Sydney

* Study Pls

Page 23