Automated Screening of Look-Alike, Sound-Alike Medicine Names for Safety

Helen Dowling
Senior Project Officer
Australian Commission on Safety and Quality in Health Care (ACSQHC)
12 August 2019

AUSTRALIAN COMMISSION ON SAFETY AND QUALITY IN HEALTH CARE

UNIVERSITY of TASMANIA AUSTRALIA

Curtin University

Acknowledgements

- Girish Swaminathan
- Diana Shipp
- Christopher Leahy

AUSTRALIAN COMMISSION ON SAFETY AND QUALITY IN HEALTH CARE

- Dr Colin Curtain, University of Tasmania

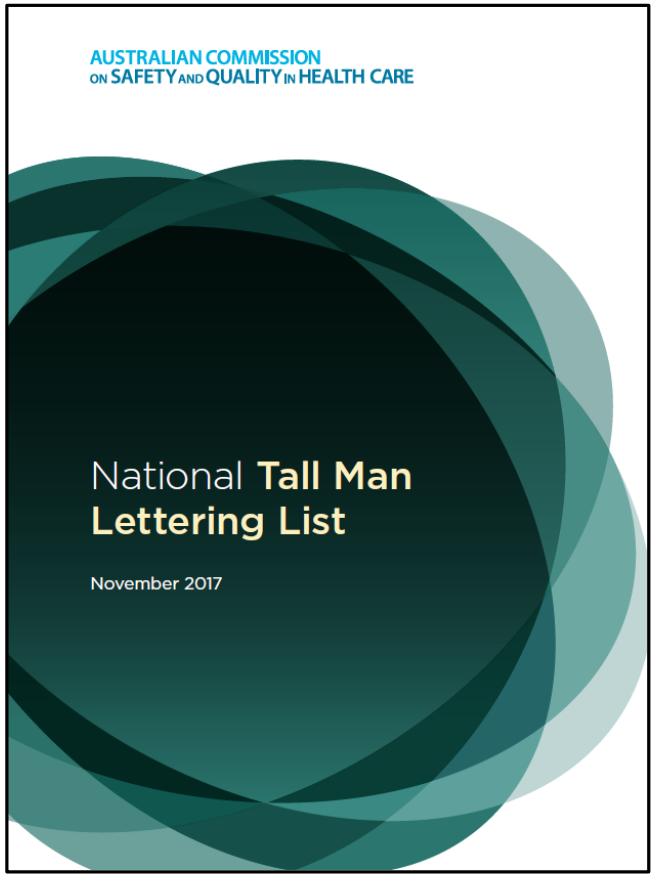
UNIVERSITY of TASMANIA AUSTRALIA

- Professor Lynne Emmerton, Curtin University

Curtin University

Look-Alike, Sound-Alike (LASA) Medicines

primaXIN	primaCOR
	primaCIN
zinVit	zinNAt
zoCOR	zoTON
zoLOFT	zoCOR



Background

- Lists of LASA medicines are compiled:[1]
  - Reactively:
    - LASA lists from other countries
    - Error reports (name confusion)
    - Near-miss reports
  - Proactively:
    - Opinion surveys (potentially-confusable medicines)

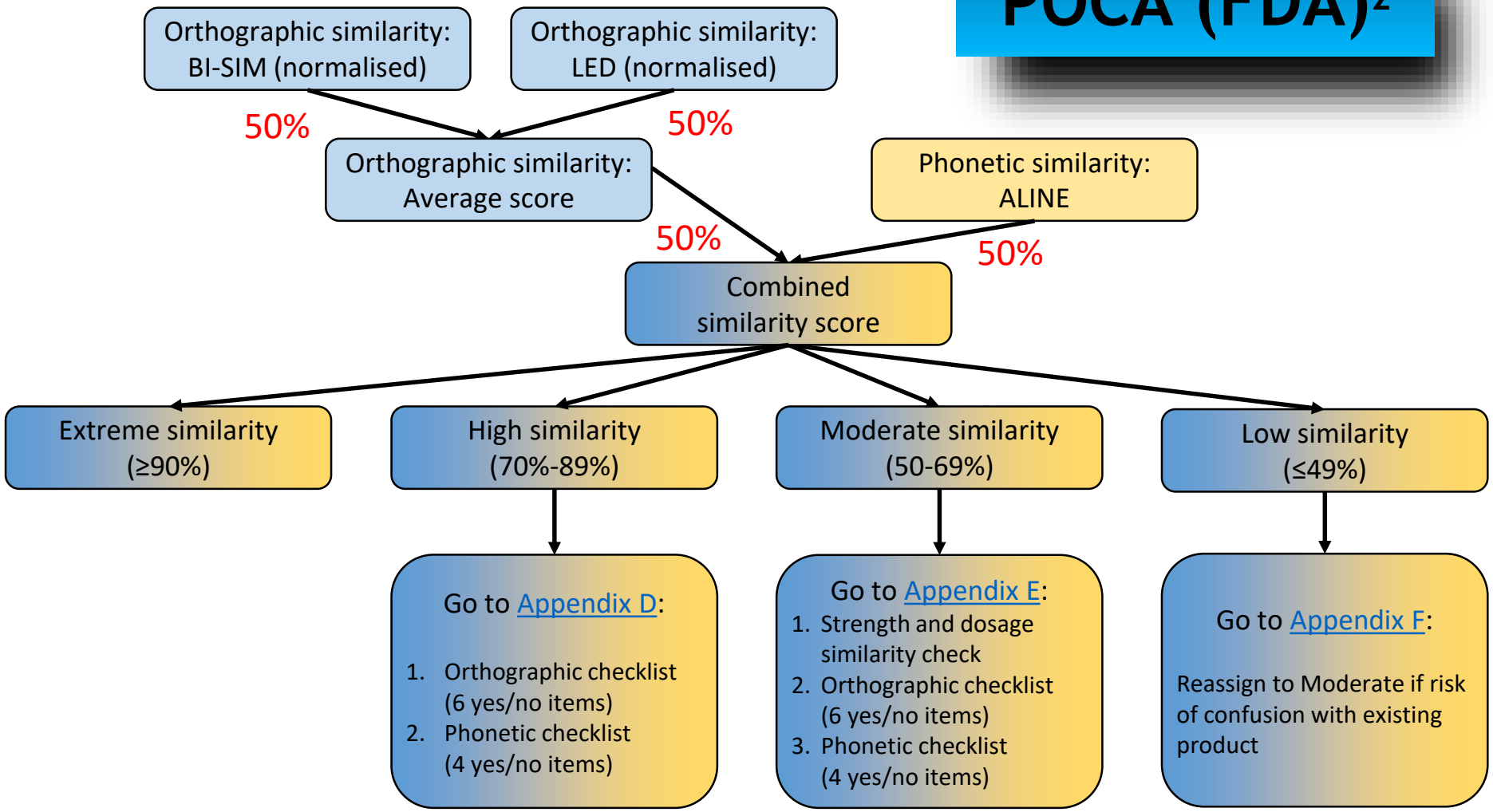
1. Emmerton L. Revision of the Tall Man lettering methodology. ACSQHC; 2016.

The Issues

- Under-reporting of errors and near misses
- Can we be more proactive, rather than reactive?
- Are there other LASA medicines that should be prioritised for the National Tall Man Lettering List?
- Could we even avert LASA medicine names from being approved in Australia?

www.tga.gov.au

POCA (FDA)[2]



2. US Food and Drug Administration. Phonetic and orthographic computer analysis (POCA) program. 2016.

Objectives

- Development phase:
  - Produce an Australian adaptation of POCA software for automated screening of LASA medicines
- Evaluation phase:
  - Review outputs
  - Compare *computed* name similarity scores to the *manually-calculated* similarity scores that underpinned the 2011 *National Tall Man Lettering List*

Methods: Development Phase

- Database: Australian Medicines Terminology
  - International spelling, e.g. amoxicillin AND amoxycillin
  - Generic and brand names
- Name transformations required
  - Brackets, numbers, slashes, hyphens deleted
  - >1 word names truncated, e.g. ‘isopto’, ‘MS’, ‘forte’, salts
- Code for BI-SIM, LED, ALINE sourced and tested
- Programmed in Python and trialled with C++
- Efficiencies and limiters reviewed

LASA v2: Look-alike sound-alike Automated Screening Application

Name 1 vs Name 2, or
Name 1 vs Tall Man List, or
Name 1 vs AMT database

MainWindow

File

Name 1: amoxicillin

Name 2:

☑ Only meeting thresholds

Orthographic % 50 Phonetic % 50

Minimum letters 5

Thresholds: Mod 0.65 High 0.69 Extreme 0.9

Go ○ Compare to Name 2 ○ Compare to tall man ◉ Compare to AMT names

Comparison of: amoxicillin to AMT names
Source converted from 1 names to 1 processed names
Target converted from 10164 names to 4528 processed names

	Source	Target	BI-SIM	LED	ALINE	COMBINED	Similarity	Source Original	Target Original
1	amoxicillin	amoxycillin	0.9091	0.9091	0.9825	0.9458	extreme		
2	amoxicillin	ampicillin	0.8182	0.8182	0.8432	0.8307	high		
3	amoxicillin	abbocillin	0.6818	0.6667	0.7387	0.7065	high		
4	amoxicillin	bicillin	0.6364	0.6364	0.7525	0.6944	high		
5	amoxicillin	aspecillin	0.6364	0.6364	0.7297	0.683	moderate		
6	amoxicillin	amoxiclav	0.6818	0.6364	0.7058	0.6824	moderate		
7	amoxicillin	dicloxacillin	0.6154	0.6154	0.7156	0.6655	moderate		

Note: In all scores, the closer to 1 represents similarity, the closer to 0 represents dissimilarity.
Current AMT version: AMT_2017_02.csv

Output:
'amoxicillin' vs AMT
(extreme → moderate)

Methods: Evaluation Phase

- Description of output file
  - Medicine name pairs with ‘moderate’, ‘high’, ‘extreme’ similarity
  - ‘Frequent flyer’ medicines
- Comparisons with ‘manual’ prioritisation of LASA name pairs (used for 2011 Tall Man List)
  - Similarity scores: manually-calculated scores mapped to computed scores
  - Risk categories: computed risk categories vs risk categories from expert consensus

Description of Output File

- Computation time for all-against-all screening >15 hours (~10 million valid comparisons)
  - Recommend periodic clean-ups rather than re-runs
- LASA pairs with a computed similarity score of at least 0.6600 ('moderate' similarity), after deletion of duplicates and self-paired names: n = 7,750
- 34 pairs with 'extreme' similarity (score ≥0.9000)

Name Pairs with ‘Extreme’ Similarity

hexachlorophane	hexachlorophene	pyrethrin	pyrethrins
diethylamine	primacin	primaxin	ennosides
dlalphatocopherol			lutose
dalphatocopherol	dalphatocopheryl	tacoccord	taxoccord
ciclosporin			
amoxicillin			
carbohydra			
triglyceride			
amoxiclav			
dalphatocop			
dalphatocop			
mitoxantron			
estradiol			
centavite			in
benzatropin			
quepine			
gentian	gentiana	cocoamphodiacetate	cocoamphodiacetic

primaCin = antimalarial primaquine
primaXin = antibacterial imipenem

Computed similarity score = 0.9034

Manually-calculated similarity score = 0.6125, but had been included in Tall Man List due to ‘extreme’ risk category (expert consensus)

Name Pairs with ‘Extreme’ Similarity

hexachlorophane	hexachlorophene	pyrethrin	pyrethrins
diethylamine			nosides
dlalphatocopherol			ose
dalphatocopherol	dalphatocopheryl	tacoccord	taxoccord
ciclosporin			
amoxicillin			
carbohydra			
triglyceride			
amoxiclav			
dalphatoco			
dalphatoco			
mitoxantror			
estradiol			
centavite			
benzatropine	benztropine	amohexal	mohexal
quepine	quipine	minomycin	mitomycin
gentian	gentiana	cocoamphodiacetate	cocoamphodiacetic

minomycin mitomycin

minomycin = antibacterial minocycline (tablet)
mitomycin = cytotoxic available (injection)

Computed similarity score = 0.9019
Not in Tall Man Lettering List

→ Consider risk of confusion in practice

‘Frequent Flyer’ Medicines: The Top 50 with Computed Scores ≥0.6600

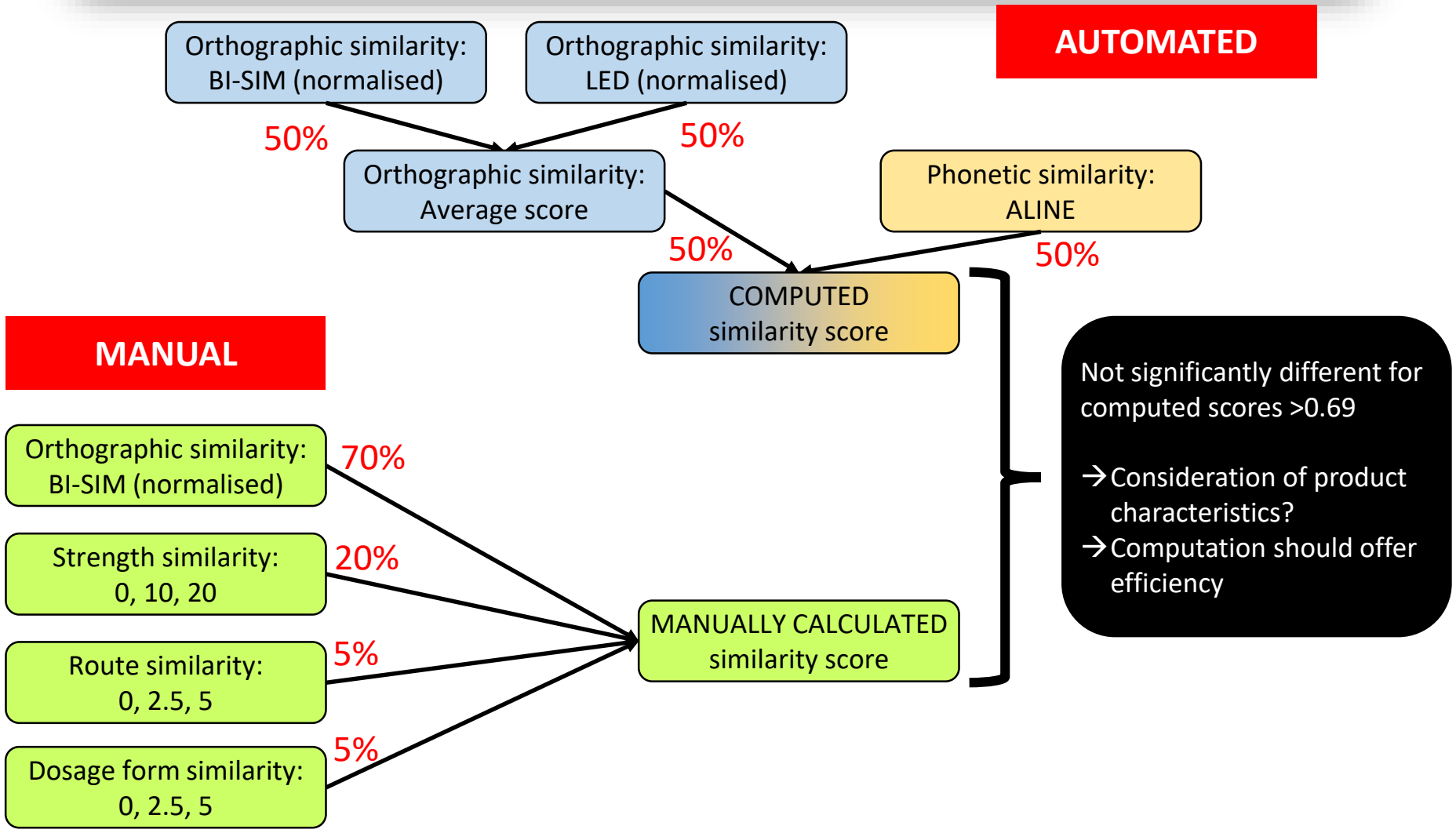
propine	39	amohexal	26	finaccord	23	baclohexal	21	clopine	20
procaine	32	acihexal	25	gabaccord	23	clopaccord	21	levohexal	20
prozine	32	anaccord	25	azahexal	22	enahexal	21	pirohexal	20
prostin	28	atropine	25	betadine	22	exaccord	21	pravaccord	20
proven	28	mohexal	25	diclohexal	22	gabahexal	21	clamohexal	19
famohexal	27	protamine	25	iprofen	22	ropaccord	21	donaccord	19
isohexal	27	protein	25	metohexal	22	zolaccord	21	gemaccord	19
pizaccord	27	proxen	25	procid	22	calamine	20	letraccord	19
proline	27	temaccord	25	prodeine	22	carbaccord	20	nifehexal	19
talohexal	27	atehexal	24	sotahexal	22	celazadine	20	parahexal	19
								ranihexal	19

Of most concern:

→ ‘pro-’ prefix

→ +/- ‘-ine’, ‘-eine’ or ‘-en’ suffix

Manual vs Computed Similarity Scores



Manual vs Computed Risk Categories

AUTOMATED

Computed similarity score

- **Extreme** similarity (≥0.9000)
- **High** similarity (0.7000-0.8900)
- **Moderate** similarity (0.6600-0.6900)
- **Low** similarity (<0.6600%)

Manual risk categories were generally one higher than the computed rating

→ Including clinical judgement **amplifies** the risk rating

MANUAL

Likelihood of Confusion	Potential Severity: Minimum	Minor	Moderate	Major	Catastrophic
1	M	H	E	E	E
2	M	H	H	E	E
3	L	M	H	H	E
4	L	M	M	H	H
5	L	L	L	M	M

Discussion

- Automation offers a proactive approach to identification of drug name similarity
- Computation time is considerable
  - Utilise one-vs-all screening option
  - Can run full periodic updates
- User interface is operational and versatile
- Interest from TGA for pre-marketing screening of proposed medicine names

Discussion

- Similarity scores have high sensitivity
  - The cost of high sensitivity is ‘noise’ in the data
- Resulting risk categories are dampened, but clinical opinion can be used in interpretation of the data
- *LASA v2* software is therefore recommended for:
  - Confirmation and updating of the Tall Man Lettering List
  - One-against-all screening of medicines in error reports
- But MUST supplement the computed scores with clinical risk considerations (indication, dosage forms, storage proximity)

Work in Progress

- Application to monoclonal antibodies (-mab) and tyrosine kinase factor inhibitors (-nib), due to risks:
  - Complicated written and spoken names e.g. daratumumab, ixekizumab
  - Alphabetical appearance in drop-down lists
  - Potency (chemotherapy)
  - Similarity in clinical use
  - Ongoing expansion of these medicine classes

→ Separate list(s) of confusable 'specialist' medicines

→ ~31 medicines grouped in pairs or trios: prioritised according to name similarity and clinical factors

Application for Health Informatics

- Work with safety and quality experts to draw attention to confusable medicines
- Be alert to environmental factors leading to confusion of medicines, e.g. alphabetical proximity in electronic lists, clicking errors
- Potential research: linkage of EMM records and clinical incident databases, with artificial intelligence to predict errors involving confusable medicines


thank you
gracias
merci
danke
obrigado
謝謝
teşekkür ederim
ngiyabonga
go raibh maith agat
mochchakkeram
sukriya
kop khun krap
terima kasih
감사합니다
dziękuję
bedankt
спасибо
tapadh leat
dakujem
arigatô
grazie
takk
dank je
мерси
ευχαριστώ
ありがとう
рахмат
Баярлалаа
faafetai lava
hvala
köszönöm
sagolun
chnorakaloutioun
gratias ago
gràcies
misaotra
matondo
paldies
grazzi
mahalo
хвала
asante
manana
obrigada
tenki
murakoze
chokrane
mamnun
дякую
djiere dieuf
tau
sulpáy
taiku
akun dankon
aciū
mersi
kia ora
barka
welalin
tack
spas
vinaka
спасибі
blagodaram
kiitos
dankie
nanni
nandri
dhanyavad
maururu
bayarlalaa
gracie
enkosi
sobodi
děkuji
mèsi
didi madloba
najis tuke
kam sah hamnida
rahmat
rahmet
tanemirt
diolch
xièxie
dhanyavadagalu
shukriya
mercé
trugarez

Safetyandquality.gov.au

Twitter.com/ACSQHC

Youtube.com/user/ACSQHC

Helen Dowling
helen.dowling@safetyandquality.gov.au
www.safetyandquality.gov.au

AUSTRALIAN COMMISSION
ON SAFETY AND QUALITY IN HEALTH CARE