Data saturation of themes – when do we stop?

Naomi Reay, Jackie Hill, Alan Tennant, Claire Hale

University of Leeds
The context of the study

Study of patient-perceived quality of life (QoL) in Systemic Sclerosis
‘Leather man disease’
Heterogeneous path
Qualitative aspect of the study

Three study aims met by mixed methodologies.

One study aim was to develop a descriptive framework of QoL in SSc was to be met using ‘one to one’ interviews. The theoretical basis was ‘needs based’ QoL; the data analysis strategy was use of thematic analysis.
Sample grid characteristics of age, disease duration, gender and disease sub-type

- 5 male, 26 Female (reflects disease distribution)
- 13 of the 31 had a disease duration of 5 years or less; 18 had SSc for over 5 years
- 16 were aged 50 or over, 15 were under 50
- 16 had diffuse disease, 15 had limited disease
The process of thematic analysis used

- Verbatim transcripts checked
- Emersion of the researcher into the data and note making
- Reflection and re-reading of transcripts – development of themes and thematic definitions. Review by second researcher
- Reflection and re-reading of transcripts – refinement of thematic definitions and thematic coding within the text. Review by second researcher
The Descriptive Framework
Data saturation

- Data saturation is the term applied to the point at which no new data emerge

Data saturation

- 30 themes identified
- On revisiting the interview transcripts all themes were represented within the first four interviews. Not all sub-themes were present
- How does that compare with others – 12
- Other uses of the interview data – instrument development which drew items from across the range of interviews
When do we stop?

- Determining sample size is ultimately a matter of judgement and experience in evaluating the quality of the information collected against the uses to which it will be put, the particular research method and purposeful sampling strategy employed....

Sandelowski (1995) Sample size in qualitative research
Research in Nursing and Health 18 (2): 179-83.
Questions generated ..

- Is data saturation be met by presence of themes or does it require sub-themes?
- What factors influence data saturation –
  - Participants/Homogeneity – would any four have produced the same?
  - Researcher skills
  - Theoretical basis of interviews
  - The process of thematic analysis
  - Order of interviews
- What are the implications of this for sample size/characteristics