INTERVIEW PREPARATION · CLINICAL STATISTICAL PROGRAMMING

From Endpoint to TLF

A principal-level narrative playbook. Three therapeutic areas — oncology, CNS and psychiatry — each walked end to end from the protocol objective through SDTM and ADaM to the procedures, the tables, the figures and the read-out.

Prepared for	Bhanoji Duppada
Target roles	Principal / Associate Director – Statistical Programming
Contents	3 full talk tracks, 5 modules, 30-question panel drill
Use	Read the talk tracks aloud, timed
Built	05 August 2026

How to use this pack

At principal and associate-director level, nobody wants to hear that you "have experience in SDTM, ADaM and TLF." Every candidate says that. What separates you in the room is whether you can take a **PROTOCOL OBJECTIVE** and walk it, without stopping, all the way down to the **column in the dataset**, the **procedure that consumed it**, the **table number it landed in**, and finally **whether the study read out positive or not** — and then walk back up and explain what you would have done differently.

That walk is a single continuous chain. This pack teaches the chain three times, in three therapeutic areas, so that whichever one the panel names you have a rehearsed forty-minute route through it.

The spine — seven links, always in this order

Every narrative in this pack follows the same seven links. Learn the spine once and the therapeutic area becomes interchangeable detail.

#	Link	The sentence you say
1	**Objective**	"The primary objective was to demonstrate …"
2	**Endpoint + estimand**	"The primary endpoint was X, and the estimand treated discontinuation as …"
3	**Collection**	"That came off the CRF as … , collected at these visits."
4	**SDTM**	"Which lands in domain … , with these variables carrying the signal."
5	**ADaM**	"Which I derived into dataset … , PARAMCD = … , AVAL = … , with these flags."
6	**Analysis**	"The SAP called for … , which is PROC … with these options."
7	**Output + read-out**	"That produced table … and figure … . The result was … , and here is what it meant."

> **THE SINGLE MOST IMPORTANT HABIT.** Never say a dataset name without immediately saying the variable that carries the endpoint. "ADTTE" means nothing. "ADTTE, where PARAMCD is OS, AVAL is months from randomisation, and CNSR is zero for a death and one for a censor" is the answer of someone who has actually built it.

The four blocks in every section

Block	What it is for
Say this	The words, at interview pace. Read aloud until it is fluent.
Derivation	The actual logic. What a follow-up question will demand.
Code	Runnable SAS, so you can defend the options, not just name the procedure.
Trap	The thing that catches people at this level. Panels probe here.

The forty-minute route

You will rarely be given forty uninterrupted minutes. You will be given eight, and if you are good, the panel keeps pulling. Structure your material so that any link in the spine can be expanded on demand without losing your place.

Minutes	What you are doing	Where you are in the spine
0–3	Set the study: phase, population, arms, N, duration, who the sponsor was	Context
3–8	Objectives and endpoints, primary then key secondary, with the estimand	1–2
8–15	Data flow: CRF → SDTM domains, and the two or three that carry the endpoint	3–4
15–25	ADaM: dataset by dataset, PARAMCD by PARAMCD, the derivations you owned	5
25–33	The analysis: procedures, model statements, why those options	6
33–40	The TLF inventory, the figures, and the read-out	7
Then	“Where it went wrong, and what I changed” — the part that gets you hired	Reflection

Three studies, one shape

	Oncology	CNS (Alzheimer's)	Psychiatry (MDD)
Primary endpoint type	Time-to-event	Continuous, repeated measures	Continuous, repeated measures
Primary ADaM	ADTTE	ADQS / ADEFF	ADQS / ADEFF
Core SDTM	TU, TR, RS, DS, DM	QS, plus imaging and biomarker	QS
Primary procedure	LIFETEST, PHREG	MIXED (MMRM)	MIXED (MMRM)
Signature figure	Kaplan–Meier	LS mean change over time	Response and remission over time
Signature secondary	ORR, DOR	Responder / time-to-worsening	Response ≥50%, remission
Where it usually fails	Censoring rules	Missing data at 78 weeks	Placebo response

If you learn only one thing from the table above: **THE SHAPE OF THE ANALYSIS FOLLOWS THE SHAPE OF THE ENDPOINT, NOT THE DISEASE.** A time-to-event endpoint in psychiatry (time to relapse) is analysed exactly like overall survival. Say that out loud in an interview and you sound like someone who thinks in structures rather than in job titles.

A caution on how you describe your own work

Be precise about what you personally owned. "I built the ADTTE and the survival outputs, and I was the independent QC programmer on the response datasets" is stronger and safer than an implied claim to have run the whole study. Panels at this level check, and the person who over-claims once loses the benefit of the doubt on everything else they said.

Oncology — the survival narrative

The study you describe

SAY THIS. "The one I usually walk people through is a Phase 3, randomised, open-label study in previously untreated advanced non-small-cell lung cancer. Roughly six hundred subjects, randomised one-to-one to the investigational agent plus platinum doublet, or platinum doublet alone, stratified by PD-L1 expression, histology and region. Treatment until progression or unacceptable toxicity, with survival follow-up every twelve weeks thereafter. Two interim analyses and a final, with an alpha-spending function."

Give the panel that in twenty seconds. It buys you the right to talk about everything else, because now they can picture the data.

Objectives and endpoints

	Endpoint	Analysis population	Assessed by
Primary	Overall survival (OS)	ITT	Death from any cause
Key secondary	Progression-free survival (PFS)	ITT	BICR per RECIST 1.1
Key secondary	Objective response rate (ORR)	ITT	BICR per RECIST 1.1
Secondary	Duration of response (DOR)	Responders	BICR
Secondary	Disease control rate (DCR)	ITT	BICR
Secondary	Time to response (TTR)	Responders	BICR
Safety	TEAEs, grade ≥3, SAEs, discontinuations	Safety	CTCAE v5.0
Exploratory	PFS2, PRO by EORTC QLQ-C30	ITT	—

TRAP. If you say "PFS" without saying who assessed it, expect the follow-up immediately. Investigator-assessed and blinded independent central review give different datasets, different RS records, and in an open-label study the difference is the entire credibility of the endpoint. You carry both, you analyse BICR as primary, and you present investigator assessment as a supportive concordance analysis.

Estimands — say this before they ask

SAY THIS. "The primary estimand for OS was a treatment-policy strategy. Overall survival is robust to intercurrent events precisely because you keep following the subject regardless of what they do next — discontinuation, crossover, subsequent therapy. So the estimand was: difference in survival distribution between randomised arms, in all randomised subjects, regardless of adherence or subsequent anti-cancer therapy. For PFS we used the same treatment-policy framing for the primary and then a set of censoring-rule sensitivity analyses, because PFS is not robust in the same way — if a subject starts a new therapy before documented progression, the observed progression is no longer attributable to the randomised treatment."

That paragraph is the single highest-yield thing in this module. It shows you understand ICH E9(R1) as an operating principle rather than a compliance sentence.

SDTM — where an oncology endpoint actually lives

Domain	What it carries	The variables that matter
DM	Demographics, and the death date	`DTHDTC`, `DTHFL`, `RFSTDTC`, `ARM`, `ACTARM`
DS	Disposition, including death and study discontinuation	`DSDECOD`, `DSTERM`, `DSSTDTC`, `DSCAT`
TU	Tumour identification — which lesion is which	`TULNKID`, `TULOC`, `TUORRES` (TARGET / NON-TARGET / NEW), `TUEVAL`
TR	Tumour results — the measurements	`TRLNKID`, `TRTESTCD` (`LDIAM`, `SUMDIAM`), `TRSTRESN`, `TRDTC`, `TREVAL`
RS	Disease response at each assessment	`RSTESTCD` (`OVRLRESP`, `BESTRESP`), `RSORRES` (CR/PR/SD/PD/NE), `RSDTC`, `RSEVAL`, `RSEVALID`
EX	Exposure	`EXTRT`, `EXDOSE`, `EXSTDTC`, `EXENDTC`
CM	Concomitant meds, **including subsequent anti-cancer therapy**	`CMDECOD`, `CMSTDTC`, `CMINDC`
PR	Procedures — subsequent radiotherapy, surgery	`PRTRT`, `PRSTDTC`
AE, LB, VS	Safety	standard
SV / SU	Visit occurrence, survival follow-up contacts	`SVSTDTC`

TRAP. The question that catches people: "Where does the last-known-alive date come from?" It is not one domain. It is the maximum of every date on which the subject was demonstrably alive — last visit in SV, last dose in EX, last assessment in RS/TR, last AE start, last lab collection, last survival-follow-up contact. If you censor OS at the last tumour assessment you have thrown away months of follow-up and biased the analysis. Building `LSTALVDT` correctly in ADSL is a genuine piece of programming and worth two minutes of airtime.

TRAP. `RSEVAL` and `RSEVALID` are how you separate the assessors. `RSEVAL = 'INDEPENDENT ASSESSOR'` with `RSEVALID = 'RADIOLOGIST 1'` / `'RADIOLOGIST 2'` / `'ADJUDICATOR'` versus `RSEVAL = 'INVESTIGATOR'`. Your ADaM selects on these. Getting this wrong silently mixes two assessment streams into one endpoint.

ADSL — the variables the whole study leans on

```
/* Last known alive date: the maximum over every domain that proves the
   subject was alive on that date. Missing here costs you follow-up time. */
data lstalv;
  set  ex   (keep=usubjid exstdtc exendtc  rename=(exendtc=dt))
       sv   (keep=usubjid svstdtc          rename=(svstdtc=dt))
       rs   (keep=usubjid rsdtc            rename=(rsdtc=dt))
       tr   (keep=usubjid trdtc            rename=(trdtc=dt))
       lb   (keep=usubjid lbdtc            rename=(lbdtc=dt))
       ae   (keep=usubjid aestdtc          rename=(aestdtc=dt));
  if length(dt) >= 10;                /* complete dates only */
  _d = input(substr(dt,1,10), yymmdd10.);
run;

proc sql;
  create table adsl_lst as
  select usubjid, max(_d) as LSTALVDT format=date9.
  from lstalv group by usubjid;
quit;
```

Key ADSL variables for this study: `RANDDT`, `TRTSDT`, `TRTEDT`, `DTHDT`, `DTHFL`, `LSTALVDT`, `ITTFL`, `SAFFL`, `RESPFL`, `PDL1GR1` (stratum as randomised) and `PDL1GR2` (stratum as collected), `STRAT1`–`STRAT3`, `TRT01P`, `TRT01A`, `AGEGR1`, `SUBSQTFL`, `SUBSQTDT`.

TRAP. Randomised strata versus actual strata. Subjects get randomised on the stratum value entered in IRT, which is sometimes wrong. The SAP will say the primary analysis uses the **stratum as randomised** and a sensitivity analysis uses the corrected value. Carry both in ADSL. Say this unprompted — it signals you have lived through a stratification discrepancy memo.

ADTTE — the dataset that carries the primary endpoint

ADTTE is a BDS dataset, one record per subject per parameter. The structure is small and every variable earns its place.

Variable	Meaning	Example (OS, a death)	Example (OS, censored)
`PARAMCD`	Parameter	`OS`	`OS`
`PARAM`	Label	Overall Survival (months)	Overall Survival (months)
`STARTDT`	Time origin	Randomisation date	Randomisation date
`ADT`	Event or censoring date	Death date	Last known alive date
`AVAL`	Time	`(ADT-STARTDT+1)/30.4375`	same formula
`CNSR`	0 = event, 1 = censored	`0`	`1`
`EVNTDESC`	What the event was	`Death`	—
`CNSRDTDSC`	Why censored	—	`Last known alive`
`SRCDOM` `SRCVAR` `SRCSEQ`	Traceability to SDTM	`DM` / `DTHDTC` / .	`ADSL` / `LSTALVDT` / .

TRAP — THE ONE THAT SEPARATES LEVELS. `CNSR = 0` means **event**. It is an indicator of censoring, so zero means not censored. Half of candidates get this backwards under pressure, and it inverts every hazard ratio. Also note that `CNSR` is CDISC's convention while `PROC PHREG` and `PROC LIFETEST` want you to declare the censoring value explicitly — hence `time AVAL*CNSR(1)`.

OS derivation

SAY THIS. “OS is time from randomisation to death from any cause. If the subject died, the event date is the death date and CNSR is zero. If not, they are censored at the last date known alive, and CNSR is one. Anyone with no post-randomisation contact at all is censored at day one on the randomisation date, so they contribute zero follow-up rather than being dropped — you never drop a randomised subject out of an ITT analysis.”

```
data adtte_os;
  merge adsl(in=a keep=usubjid studyid siteid randdt dthdt dthfl lstalvdt
                       trt01p trt01pn trt01a trt01an ittfl strat1 strat2 strat3);
  by usubjid;
  if a and ittfl='Y';

  PARAMCD='OS'; PARAM='Overall Survival (months)'; PARAMN=1;
  STARTDT = randdt;

  if dthfl='Y' and n(dthdt) then do;
      ADT=dthdt;  CNSR=0;  EVNTDESC='Death';
      SRCDOM='DM'; SRCVAR='DTHDTC';
  end;
  else do;
      ADT = max(lstalvdt, randdt);  CNSR=1;
      CNSRDTDSC='Last known alive';
      SRCDOM='ADSL'; SRCVAR='LSTALVDT';
  end;

  AVAL  = (ADT - STARTDT + 1) / 30.4375;   /* months */
  ADY   =  ADT - STARTDT + 1;
  format STARTDT ADT date9.;
run;
```

PFS derivation — where the real programming is

SAY THIS. “PFS is time from randomisation to the earlier of documented progression per RECIST 1.1 by central review, or death from any cause. Censoring is where the work is. A subject with no baseline-adequate post-baseline assessment is censored at randomisation. Otherwise censoring is at the last adequate tumour assessment. The FDA convention also censors at the last assessment before a new anti-cancer therapy, and censors progression that follows two or more consecutively missed assessments; the EMA convention is more permissive. We ran the FDA rule as primary and the others as pre-specified sensitivity analyses, and we tabulated them side by side.”

```
/* 1. earliest central-review PD */
proc sql;
  create table pd as
  select usubjid, min(ADT) as PDDT format=date9.
  from  adrs
  where PARAMCD='OVRLRESP' and AVALC='PD' and RSEVAL='INDEPENDENT ASSESSOR'
  group by usubjid;

/* 2. last adequate (non-NE) tumour assessment */
  create table lastadq as
  select usubjid, max(ADT) as LSTADQDT format=date9.
  from  adrs
  where PARAMCD='OVRLRESP' and AVALC in ('CR','PR','SD','PD')
  group by usubjid;
quit;

data adtte_pfs;
  merge adsl(in=a) pd lastadq;
  by usubjid;
  if a and ittfl='Y';
  PARAMCD='PFS'; PARAM='Progression-free Survival (months)'; PARAMN=2;
  STARTDT = randdt;

  _evdt = min(coalesce(PDDT, .z), coalesce(dthdt, .z));

  if n(_evdt) and _evdt ne .z and
     (SUBSQTDT = . or _evdt <= SUBSQTDT) then do;
      ADT=_evdt; CNSR=0;
      EVNTDESC = ifc(PDDT=_evdt,'Disease progression','Death');
  end;
  else if LSTADQDT = . then do;
      ADT=randdt; CNSR=1; CNSRDTDSC='No adequate post-baseline assessment';
  end;
  else do;
      ADT=min(LSTADQDT, coalesce(SUBSQTDT-1, LSTADQDT));
      CNSR=1; CNSRDTDSC='Last adequate tumour assessment';
  end;

  AVAL=(ADT-STARTDT+1)/30.4375;
run;
```

TRAP. Interviewers love: “A subject progresses at week 24 but their last scan before that was week 6 — the week-12 and week-18 scans were missed. What do you do?” Answer: under the FDA rule that is progression documented after two or more consecutively missed assessments, so PFS is censored at the week-6 assessment, not counted as an event at week 24. Under the EMA rule you would generally take the event. This is exactly why the sensitivity table exists.

ADRS — the response dataset

ADRS is BDS, one record per subject per parameter per analysis visit.

PARAMCD	PARAM	Structure	AVALC / AVAL
`OVRLRESP`	Overall response at each timepoint	per visit	CR / PR / SD / PD / NE, `AVAL` 1–5
`BOR`	Best overall response	one per subject	CR / PR / SD / PD / NE
`CBOR`	Confirmed best overall response	one per subject	requires confirmation ≥4 weeks later
`OBJRSP`	Objective response (responder yes/no)	one per subject	`AVAL` 1/0, `CRIT1`, `CRIT1FL`
`DISCTRL`	Disease control (CR/PR/SD ≥n weeks)	one per subject	`AVAL` 1/0

SAY THIS ON CONFIRMATION. "In RECIST 1.1 for a non-randomised or ORR-primary setting, a partial response has to be confirmed by a repeat assessment no less than four weeks later. So CBOR is not a sort of OVRLRESP — it is a derived sequence check: a PR followed by a subsequent CR or PR, with no intervening PD. Stable disease also has a minimum duration requirement measured from randomisation. Programming CBOR is a look-ahead within subject, and it is where most defects in a response dataset live."

```
/* Responder flag with the criterion recorded, per ADaM convention */
data adrs_objrsp;
  set  bor;
  PARAMCD='OBJRSP'; PARAM='Objective Response';
  CRIT1  = 'Confirmed CR or PR per RECIST 1.1 (BICR)';
  if AVALC in ('CR','PR') then do; AVAL=1; CRIT1FL='Y'; end;
  else do;                         AVAL=0; CRIT1FL='N'; end;
run;
```

TRAP. ADaM requires that when you use `CRIT1FL` you also populate `CRIT1` with the human-readable criterion. A define.xml reviewer will look for exactly that pairing.

ADTR — the dataset behind the waterfall

ADTR carries tumour burden over time: `PARAMCD = 'SUMDIAM'` (sum of diameters of target lesions), with `BASE` = baseline sum, `CHG`, `PCHG`, and a derived `NADIR` and `PCHGNAD` (percent change from nadir, which is what actually defines progression of target disease — a 20 percent increase from nadir and at least 5 mm absolute increase).

TRAP. "Why do you need both the 20 percent and the 5 mm?" Because in small-burden disease a 20 percent relative increase can be two millimetres, which is inside measurement error. The absolute floor exists to stop measurement noise being called progression. That answer demonstrates you understand the clinical intent, not just the rule.

The analysis — procedures, and why those options

Kaplan–Meier medians and the stratified log-rank

```
proc lifetest data=adtte(where=(PARAMCD='OS')) method=km
              conftype=loglog  alpha=0.05
              plots=survival(atrisk=0 to 48 by 6 cb=hw test);
  time  AVAL*CNSR(1);
  strata STRAT1 STRAT2 STRAT3 / group=TRT01PN;   /* stratified log-rank */
  ods output Quartiles=q  ProductLimitEstimates=ple  HomTests=lr;
run;
```

- `CNSR(1)` declares 1 as the censoring value — matching the CDISC convention.
- `conftype=loglog` gives the complementary log–log confidence interval, the standard choice for KM medians and landmark rates because it keeps the interval inside (0,1).
- `strata ... / group=` is what makes it a **stratified** log-rank. Putting the treatment variable on the `strata` statement alone gives you an unstratified test — a classic silent error.

Milestone rates (12- and 24-month survival) come from `ProductLimitEstimates` with Greenwood standard errors, again on the log–log scale.

Hazard ratio

```
proc phreg data=adtte(where=(PARAMCD='OS'));
  class TRT01PN(ref='0') / param=ref;
  model AVAL*CNSR(1) = TRT01PN / risklimits ties=efron;
  strata STRAT1 STRAT2 STRAT3;
  hazardratio 'Treatment vs Control' TRT01PN / diff=ref;
  ods output ParameterEstimates=pe HazardRatios=hr;
run;
```

TRAP. "Why Efron and not Breslow?" Efron handles tied event times more accurately than Breslow when ties are common, and event times in oncology tie constantly because assessments are scheduled. `ties=exact` is the most accurate but expensive. Know that the SAS default is Breslow and that your SAP should state the choice — if it does not, that is a finding you raise.

TRAP. "Your log-rank p-value and your Cox p-value disagree slightly. Which do you report?" Whichever the SAP names as primary — normally the stratified log-rank for the hypothesis test, with the Cox hazard ratio reported as the estimate of effect size. They are answering different questions and it is entirely normal for them to differ. Never retro-fit the choice to the nicer number.

ORR and the stratified difference

```
proc freq data=adrs(where=(PARAMCD='OBJRSP'));
  tables TRT01PN*AVAL / nocol nopercent
                        binomial(level='1' exact) alpha=0.05;
  exact binomial;
run;

/* stratified comparison across arms */
proc freq data=adrs(where=(PARAMCD='OBJRSP'));
  tables STRAT1*STRAT2*STRAT3*TRT01PN*AVAL / cmh;
run;
```

Within-arm ORR gets an exact Clopper–Pearson interval; the between-arm comparison gets a CMH test stratified by the randomisation factors, and the difference in rates gets a stratified Newcombe or Mantel–Haenszel weighted interval depending on what the SAP specifies.

DOR — the subtlety worth naming

DOR is time from **FIRST DOCUMENTED RESPONSE** to progression or death, analysed only in responders.

TRAP. "Why do you not report a p-value for DOR?" Because the responder subset is not a randomised comparison — you have conditioned on a post-randomisation outcome. DOR is descriptive: median with KM confidence interval, presented by arm, without a formal test. Saying this unprompted is a strong signal.

The TLF inventory — what you actually delivered

State these as a numbered inventory. It is memorable and it sounds like a real deliverable list.

Efficacy tables

#	Table	Content
14.2.1	Analysis of Overall Survival, ITT	n events, median with 95% CI, 12/24-month rates, stratified log-rank p, HR with 95% CI
14.2.2	Analysis of PFS by BICR, ITT	same structure, plus censoring-reason breakdown
14.2.3	Best Overall Response and ORR	CR/PR/SD/PD/NE counts, ORR with exact CI, difference with CMH p
14.2.4	Duration of Response, responders	median DOR, KM CI, proportion ongoing
14.2.5	Sensitivity analyses of PFS	primary rule versus alternative censoring rules, side by side
14.2.6	Subgroup analysis of OS	HR by PD-L1, histology, region, age, sex, smoking, ECOG
14.2.7	Disease control rate and time to response	supportive

Safety tables

Overall summary of TEAEs; TEAEs by SOC and PT; TEAEs by maximum CTCAE grade; grade ≥3 TEAEs; treatment-related TEAEs; SAEs; TEAEs leading to discontinuation, interruption, dose reduction; deaths and cause of death; adverse events of special interest — immune-mediated events by category; exposure and dose intensity; laboratory shift tables by CTCAE grade; haematology and chemistry over time; ECG and vital signs.

Figures

#	Figure	Procedure
14.2.1	Kaplan–Meier curve, OS, with risk table	`PROC LIFETEST` plots=survival, or `PROC SGPLOT` from `ProductLimitEstimates`
14.2.2	Kaplan–Meier curve, PFS	as above
14.2.3	Forest plot of OS hazard ratios by subgroup	`PROC SGPLOT` scatter + highlow, log-scale x axis
14.2.4	Waterfall plot, best percentage change in sum of diameters	`PROC SGPLOT` vbar, sorted descending, reference lines at +20 and −30
14.2.5	Spider plot, percentage change over time by subject	`PROC SGPLOT` series, group=usubjid
14.2.6	Swimmer plot, responders, duration and ongoing status	`PROC SGPLOT` highlow + scatter markers
14.2.7	Kaplan–Meier curve, DOR	`PROC LIFETEST`

SAY THIS ABOUT THE RISK TABLE. "The number-at-risk row underneath a KM curve is not decoration — it is how a reviewer judges whether the tail of the curve is interpretable. When only eleven subjects remain at risk at month thirty-six, the separation out there means very little, and I would say so in the interpretation rather than let the picture over-claim."

That sentence does more for you than any amount of listing procedures.

Listings

Subject disposition and analysis populations; deaths; serious adverse events; adverse events leading to discontinuation; tumour assessment and response by subject; target and non-target lesion measurements; subsequent anti-cancer therapy; protocol deviations; dose modifications.

The read-out — how you close the story

SAY THIS. "The study met its primary endpoint. Median OS was 19.4 months versus 14.1, hazard ratio 0.72, 95 percent CI 0.60 to 0.87, stratified log-rank p equal to 0.0006, crossing the pre-specified boundary at the second interim. PFS was consistent at 0.65 and ORR was 48 percent versus 32 percent. What I would flag is that the subsequent-therapy rate was materially higher in the control arm, which dilutes the OS effect rather than inflating it, so the treatment-policy estimand here is conservative. And the curves separate at around four months, not immediately — consistent with the mechanism, but it means proportional hazards is worth checking rather than assuming."

Then, unprompted:

SAY THIS. "The thing I would do differently: our censoring-rule sensitivity analyses were specified late, during the SAP amendment before unblinding, and reconciling them against the primary took a fortnight of avoidable rework. On the next study I built the censoring rules as a parameterised macro from the start, so switching the rule is an argument rather than a rewrite. That is the change I actually carried forward."

Ending on a real, small, specific process improvement is what a director-level panel is listening for. It shows you extract institutional learning from a study rather than just delivering it.

Rapid-fire drill — oncology

1. What is the time origin for OS, and why randomisation rather than first dose? — Randomisation, to preserve the randomised comparison; first dose would exclude subjects who never dosed and break ITT.
2. A subject is lost to follow-up at month 8 and the database locks at month 30. — Censored at month 8 at their last known alive date, CNSR=1.
3. Subject dies from a car accident. — Still an OS event. OS is death from **any** cause. That is what makes it objective.
4. Difference between TTP and PFS? — TTP counts progression only; deaths are censored. PFS counts both. TTP is almost never the primary.
5. What is PFS2? — Time to progression on the next line of therapy or death; an exploratory check that a PFS benefit is not simply borrowed from the subsequent line.
6. Why is `AVAL` in months? — Convention and readability; the SAP fixes the divisor. Use 30.4375 (365.25/12) and state it in the define, or use days and convert in the output.
7. How do you QC an ADTTE? — Independent re-derivation from SDTM, then a record-level compare on `AVAL`, `CNSR`, `ADT`, `EVNTDESC` and `CNSRDTDSC`. Compare the counts of events and censors by arm before you compare anything else; a difference of one there usually explains everything downstream.
8. How would you check the proportional hazards assumption? — Log-negative-log survival plots, Schoenfeld residuals via `assess ph / resample` in `PROC PHREG`, and if it fails, pre-specified alternatives such as restricted mean survival time.

CNS — the repeated-measures narrative

The study you describe

SAY THIS. "The CNS programme I take people through is a Phase 3, randomised, double-blind, placebo-controlled study of a monoclonal antibody in early symptomatic Alzheimer's disease — mild cognitive impairment due to AD and mild AD dementia, amyloid-confirmed at screening. Around seventeen hundred subjects, one-to-one, intravenous infusion every four weeks for seventy-eight weeks, stratified by APOE ε4 carrier status and by whether the subject was on a symptomatic AD medication at baseline. Primary readout at week seventy-eight, with an open-label extension after."

Note what that sentence establishes without you saying it: you know the disease stage matters, you know amyloid confirmation is an entry criterion, you know APOE ε4 is the stratifier everyone asks about, and you know the treatment period is long — which is the whole source of the analytical difficulty.

Objectives and endpoints

	Endpoint	Instrument	Direction
Primary	Change from baseline in CDR-SB at week 78	Clinical Dementia Rating – Sum of Boxes, 0–18	Lower is better
Key secondary	Change from baseline in ADAS-Cog13	0–85	Lower is better
Key secondary	Change from baseline in ADCS-ADL-MCI	0–53	**Higher** is better
Key secondary	Change in amyloid PET Centiloid	quantitative imaging	Lower is better
Secondary	Change in MMSE, iADRS composite	0–30	Higher is better
Secondary	Time to clinical worsening (CDR-SB ≥1 point, confirmed)	derived	—
Safety	ARIA-E and ARIA-H, infusion reactions	MRI-detected AESI	—

TRAP — AND THIS ONE REALLY DOES CATCH PEOPLE. The scales do not all point the same way. CDR-SB and ADAS-Cog go **up** with worsening; ADCS-ADL and MMSE go **down** with worsening. A treatment benefit is therefore a negative LS mean difference on CDR-SB and a positive one on ADCS-ADL. Your ADaM needs an explicit direction indicator so the TLF and the figure annotation cannot get it backwards. In practice I carry a `PARAM`-level attribute — often via `PARCAT2` or a metadata lookup — that says which direction favours treatment, and the output macro reads it rather than the programmer remembering.

Estimand

SAY THIS. "The primary estimand was a treatment-policy strategy for discontinuation of study drug — subjects were asked to continue assessments regardless — and a hypothetical strategy would have

been the alternative. The population was all randomised subjects with a baseline and at least one post-baseline CDR-SB. Intercurrent events we had to define explicitly were: discontinuation of study drug, initiation or dose change of a symptomatic AD medication, and death. The endpoint was the difference between arms in mean change from baseline at week seventy-eight, and the summary measure was the least-squares mean difference from the MMRM."

Where CNS data actually lives in SDTM

This is the part CNS interviews test hardest, because the instruments are messier than tumour data.

Domain	What it carries
QS	Questionnaires — CDR, ADAS-Cog, MMSE, ADCS-ADL, NPI, GDS
FT	Functional tests — some cognitive performance tests belong here rather than QS
RS	Clinical classification / response scales when the instrument is an assessor's judgement
LB	Central-lab biomarkers — plasma p-tau181, p-tau217, Aβ42/40, NfL
MI / custom findings	Imaging quantification (PET Centiloid, MRI volumetrics) — confirm your standards team's mapping; this is not uniformly implemented across sponsors
AE	ARIA-E and ARIA-H, captured as adverse events with AESI flags
CM	Symptomatic AD medications — donepezil, memantine — an intercurrent event source
EX, DS, DM, VS	standard

The QS variables that carry everything:

Variable	Role
`QSCAT`	The instrument — `CDR`, `ADAS-COG`, `MMSE`, `ADCS-ADL`
`QSSCAT`	Sub-scale or section within the instrument
`QSTESTCD` / `QSTEST`	The individual item, or the derived total when captured as a total
`QSORRES`	Original result as collected — often the text of the option chosen
`QSSTRESC` / `QSSTRESN`	Standardised, and the numeric score you actually analyse
`QSBLFL`	Baseline flag
`QSDTC`, `QSDY`, `VISIT`, `VISITNUM`	Timing
`QSEVAL`	Who assessed — matters for informant-based instruments like CDR and ADCS-ADL

SAY THIS — IT IS A GENUINELY SENIOR OBSERVATION. "CDR and ADCS-ADL are informant-based. The score depends on a study partner, and study partners change during a seventy-eight-week study. So `QSEVAL`

and the study-partner continuity flag are not administrative — a change of informant is a real source of measurement noise, and in the studies I worked on we carried a partner-change flag into ADSL and ran a sensitivity analysis on it."

TRAP. "Where does the total score come from — do you sum it in ADaM or take it from QS?" Both happen and you must know which you have. If the EDC calculated the total it arrives in QS as its own `QSTESTCD` record and you must still independently re-derive it in QC. If only items arrive, ADaM sums them, and then you own the scoring rules — including the handling of missing items, which is instrument-specific and stated in the scoring manual, not invented by the programmer.

ADaM — ADQS and the BDS structure

ADQS is a BDS dataset: one record per subject, per parameter, per analysis visit.

Variable	Content
`PARAMCD` / `PARAM`	`CDRSB` — CDR Sum of Boxes (0-18); `ADAS13`; `ADCSADL`; `MMSE`; `IADRS`
`PARCAT1`	`EFFICACY`; `PARCAT2` may carry the instrument family
`AVAL`	The total score at that visit
`BASE`	Baseline value of the same parameter
`CHG`	`AVAL - BASE`
`PCHG`	Percentage change, where clinically meaningful (rarely for CDR-SB)
`AVISIT` / `AVISITN`	Analysis visit — the windowed visit, not the raw one
`ADT`, `ADY`	Analysis date and study day
`ANL01FL`	The record selected for the primary analysis
`DTYPE`	Populated when the record is imputed or derived, e.g. `LOCF`
`AVALCAT1`	Category, where a responder definition applies
`CRIT1` / `CRIT1FL`	The criterion and its flag for responder parameters
`SRCDOM`, `SRCVAR`, `SRCSEQ`	Traceability back to QS

Visit windowing — the thing everyone under-explains

SAY THIS. "Subjects do not attend on the nominal day. Week seventy-eight might be day 520 or day 560. The SAP defines analysis windows — target day plus or minus a range — and `AVISIT` is the window a record falls into. If two records fall into the same window you need a documented tie-break rule: normally the one closest to the target day, and if two are equidistant, the later one. That rule has to be in the SAP, not decided by the programmer, and `ANL01FL` marks the record that won."

```
/* windowing with an explicit, documented tie-break */
data adqs_win;
  set adqs_raw;
  length AVISIT $40;
  select;
    when (ADY  <   1)               do; AVISIT='Baseline';  AVISITN=0;   _tgt=0;   end;
    when (1  <= ADY <= 126)         do; AVISIT='Week 12';   AVISITN=12;  _tgt=85;  end;
    when (127 <= ADY <= 294)        do; AVISIT='Week 26';   AVISITN=26;  _tgt=183; end;
    when (295 <= ADY <= 434)        do; AVISIT='Week 52';   AVISITN=52;  _tgt=365; end;
    when (435 <= ADY <= 700)        do; AVISIT='Week 78';   AVISITN=78;  _tgt=546; end;
    otherwise                       do; AVISIT='Unscheduled'; AVISITN=999;         end;
  end;
  _dist = abs(ADY - _tgt);
run;

proc sort data=adqs_win; by usubjid paramcd avisitn _dist descending ady; run;

data adqs;
  set adqs_win;
  by usubjid paramcd avisitn;
  if first.avisitn and AVISITN ne 999 then ANL01FL='Y';
run;
```

TRAP. Windows must not overlap and must not leave gaps, and every post-baseline record must land somewhere — even if that somewhere is "Unscheduled" and excluded from the primary. Run a check that counts records by `AVISIT` and confirms the total equals the input record count. If it does not, you have silently dropped data.

The primary analysis — MMRM

SAY THIS. "The primary analysis was a mixed model for repeated measures. Change from baseline as the response, fixed effects for treatment, visit, treatment-by-visit interaction, baseline score and baseline-by-visit interaction, plus the randomisation strata. An unstructured covariance matrix for the within-subject correlation, and Kenward-Roger for the denominator degrees of freedom. The treatment effect is the least-squares mean difference at week seventy-eight, taken from the interaction slice."

```
proc mixed data=adqs(where=(PARAMCD='CDRSB' and ANL01FL='Y' and AVISITN>0
                            and ITTFL='Y'))
           method=reml;
  class USUBJID TRT01PN(ref='0') AVISITN APOEFL SYMPMEDFL;
  model CHG = TRT01PN AVISITN TRT01PN*AVISITN
              BASE   BASE*AVISITN
              APOEFL SYMPMEDFL / ddfm=kr;
  repeated AVISITN / subject=USUBJID type=un;
  lsmeans TRT01PN*AVISITN / cl diff alpha=0.05;
  ods output LSMeans=lsm  Diffs=dif;
run;
```

Defend every option:

- `type=un` — unstructured. It makes no assumption about how correlation decays over visits. It is the default choice in confirmatory work because it is the least restrictive. It also costs the most parameters, so it can fail to converge.
- **The convergence fallback must be pre-specified.** "If the unstructured model fails to converge we step down through a pre-specified sequence — Toeplitz, then heterogeneous AR(1), then AR(1), then

compound symmetry — and we report which one was used." Interviewers ask this constantly and the wrong answer is "we would try things until it converged."

- `ddfm=kr` — Kenward-Roger. Adjusts the degrees of freedom and inflates the standard errors to account for estimating the covariance matrix. Without it, with unstructured covariance and moderate N, your type-I error runs high.
- `repeated` **versus** `random`**.** Here the subject effect is modelled through the residual covariance structure, not as a random intercept. If you say "random intercept" for an MMRM you have described a different model — a random-coefficients model — and a statistician on the panel will notice.
- **Why MMRM and not LOCF ANCOVA?** MMRM uses all available observations and is valid under a missing-at-random assumption, without fabricating a value. LOCF assumes the disease stopped progressing at dropout, which in a progressive neurodegenerative disease is not conservative — it is anti-conservative for a drug that slows decline, because it freezes the placebo arm's decline too.

That last point is a strong, specific, senior answer. Say it.

Missing data and sensitivity

SAY THIS. "Seventy-eight weeks in a dementia population means real dropout — ten to twenty percent, and it is not random, because the subjects who decline fastest are the most likely to leave. So the missing-at-random assumption behind MMRM is an assumption, not a fact. We ran control- based multiple imputation as the main sensitivity analysis — imputing the treated arm's missing data from the placebo arm's distribution, a jump-to-reference approach — and then a tipping-point analysis to find how extreme the unobserved data would have to be to overturn the result."

```
proc mi data=adqs_wide seed=20260805 nimpute=100 out=mi_out;
  class TRT01PN;
  var BASE CHG12 CHG26 CHG52 CHG78;
  monotone reg;
  mnar model(CHG78 / modelobs=(TRT01PN='0'));   /* control-based */
run;

proc mixed data=mi_out; by _Imputation_;  /* analyse each */ ... run;

proc mianalyze parms=lsm_by_imp;          /* Rubin's rules */
  modeleffects Estimate;  stderr StdErr;
run;
```

TRAP. "How many imputations?" Enough that the Monte Carlo error is small relative to the standard error — historically five was cited, but with modern computing 100 or more is standard, and the SAP fixes the seed so the result is reproducible. Reproducibility of the seed is a GxP point, not a statistical one, and mentioning it makes you sound like someone who has been audited.

Secondary and responder analyses

TIME TO CLINICAL WORSENING. Defined as the first visit with a CDR-SB increase of at least one point from baseline, confirmed at the next visit. This converts a continuous endpoint into a time-to-event endpoint — and now you are back in ADTTE with `PROC LIFETEST` and `PROC PHREG`, exactly as in oncology.

SAY THIS. "It is the same machinery. Once I had built ADTTE for the oncology programme, the time-to-worsening endpoint in the CNS study reused the same censoring and derivation macro with a

different event definition. That is the argument for building endpoint derivations as parameterised components rather than per-study programs."

That connects your two stories into one platform story, which is exactly the kind of thing an associate-director panel is listening for.

The TLF inventory — CNS

Efficacy tables

#	Table
14.2.1	MMRM analysis of change from baseline in CDR-SB by visit — LS means, LS mean difference, 95% CI, p-value
14.2.2	Same for ADAS-Cog13
14.2.3	Same for ADCS-ADL-MCI
14.2.4	Change from baseline in amyloid PET Centiloid, and proportion achieving amyloid negativity
14.2.5	Time to confirmed clinical worsening — KM medians, hazard ratio
14.2.6	Sensitivity: control-based multiple imputation and tipping point
14.2.7	Subgroup analyses of the primary — APOE ε4 status, baseline severity, age, sex
14.2.8	Plasma biomarker change from baseline — p-tau181, p-tau217, NfL

Safety tables — the CNS-specific ones

Overview of TEAEs; TEAEs by SOC and PT; **ARIA-E BY SEVERITY, TIMING AND RADIOGRAPHIC CHARACTERISTICS**; **ARIA-H — microhaemorrhage and superficial siderosis — by count category**; ARIA by APOE ε4 genotype (this is the table everyone in the field reads first); infusion-related reactions; ARIA leading to treatment interruption or discontinuation; MRI findings over time; symptomatic versus asymptomatic ARIA; C-SSRS summary; falls; weight.

Figures

#	Figure
14.2.1	LS mean change from baseline in CDR-SB over time by arm, with standard-error bars
14.2.2	Same for ADAS-Cog13 and ADCS-ADL
14.2.3	Cumulative distribution function of change from baseline at week 78 — the whole distribution, not just the mean
14.2.4	Forest plot of subgroup LS mean differences
14.2.5	Kaplan–Meier for time to confirmed clinical worsening
14.2.6	Mean amyloid PET Centiloid over time by arm
14.2.7	ARIA-E incidence by APOE ε4 genotype — grouped bar

SAY THIS ABOUT THE CDF PLOT. "The CDF plot is the one I would defend hardest. A mean difference of 0.45 points on CDR-SB is easy to dismiss. The cumulative distribution shows whether the whole distribution shifted or a subgroup drove it, and it is what a reviewer and an advisory committee will actually interrogate. I would rather hand over a figure that could embarrass the result than one that only flatters it."

Listings

Individual subject scores by visit; subjects excluded from the primary analysis with reason; ARIA events with MRI dates and severity; study-partner changes; symptomatic AD medication changes; protocol deviations affecting efficacy evaluability.

The read-out

SAY THIS. "The primary endpoint was met but the effect was modest — an LS mean difference on CDR-SB of around minus 0.45 points at week seventy-eight, roughly a twenty-seven percent slowing of decline, p less than 0.001. Amyloid PET showed a large and unambiguous reduction. ARIA-E occurred in about twelve percent of treated subjects overall and substantially more in ε4 homozygotes. So the honest framing is: the biomarker effect is decisive, the clinical effect is real but small and its clinical meaningfulness is genuinely debated, and the safety signal is manageable but genotype- dependent. My job as the programmer was to make sure every one of those three statements was supported by an output someone could audit, including the ones that were not flattering."

The MS variant — worth knowing, in case they ask

If the panel says multiple sclerosis rather than Alzheimer's, the endpoints change and so does the statistics:

Endpoint	Analysis
Annualised relapse rate (ARR) — primary in relapsing MS	Negative binomial regression, `PROC GENMOD dist=negbin` with `offset=log(exposure_years)`
Time to 12- or 24-week confirmed disability progression (EDSS-based)	`PROC LIFETEST` / `PROC PHREG`
Number of new or enlarging T2 lesions on MRI	Negative binomial, same structure
Change in EDSS, timed 25-foot walk, 9-hole peg test	MMRM

```
proc genmod data=adrel;
  class USUBJID TRT01PN(ref='0') REGION;
  model RELNUM = TRT01PN REGION BASEDSS
        / dist=negbin link=log offset=LOGEXPYR type3;
  estimate 'Rate ratio' TRT01PN 1 -1 / exp;
run;
```

TRAP. "Why negative binomial and not Poisson?" Because relapse counts are overdispersed — the variance exceeds the mean, since relapse propensity varies between subjects. Poisson would underestimate the standard errors and give you false significance. The `offset` handles unequal follow-up time so the model estimates a rate rather than a count.

Rapid-fire drill — CNS

1. Which direction is good on CDR-SB? — Lower. Negative change from baseline. Opposite of ADCS-ADL.
2. What if a subject has a baseline but no post-baseline value? — Excluded from MMRM by construction (no `CHG` record), but stated in the analysis-set table and covered by the sensitivity analysis.
3. Why `ANL01FL` rather than just subsetting? — Traceability. The record stays in the dataset with the reason it was not selected visible; a reviewer can reproduce the selection.
4. Where does `DTYPE` get populated? — On imputed or derived records only, never on observed ones. `DTYPE='LOCF'` on a carried-forward record; blank on a real observation.
5. Total score when three items are missing? — Follow the instrument's own scoring manual. Some allow prorating up to a threshold; some make the total missing. It is never the programmer's judgement.
6. How do you handle an unscheduled visit? — It lands in a window if the SAP permits, otherwise `AVISIT='Unscheduled'` and excluded from the primary, but retained in listings and in by-visit summaries where the SAP allows.
7. What does the treatment-by-visit interaction give you? — Separate treatment effects at each visit, which is the point; without it you would be estimating one averaged effect across all visits.
8. What is `iADRS`? — An integrated composite of a cognitive and a functional scale, used to combine two domains into a single primary in some AD programmes; you derive it in ADaM from the component totals, which makes its component missingness rules a programming decision you must document.

Psychiatry — the short, noisy, high-placebo narrative

The study you describe

SAY THIS. "The psychiatry programme I use is a Phase 3, randomised, double-blind, placebo-controlled, fixed-dose study in major depressive disorder. Around four hundred and fifty subjects randomised one-to-one-to-one across two active doses and placebo, six weeks of double-blind treatment with weekly efficacy visits, then a two-week safety follow-up. Entry required a MADRS total of at least twenty-six and a CGI-S of at least four. Stratified by site and by whether the current episode was recurrent."

Then immediately establish the thing that defines psychiatry programming:

SAY THIS. "The defining feature of this therapeutic area is not the endpoint — it is the noise. Placebo response in MDD routinely runs thirty to forty percent, dropout runs twenty to thirty percent in six weeks, and the primary instrument is a clinician-rated interview, so measurement variance is high. Everything in the analysis plan exists to deal with that: the visit schedule is weekly rather than monthly, the model is MMRM rather than a single endpoint comparison, the sensitivity analyses are about missingness, and the site-level rater reliability is monitored."

Objectives and endpoints

	Endpoint	Instrument	Range	Direction
Primary	Change from baseline in MADRS total at week 6	Montgomery–Åsberg Depression Rating Scale, 10 items × 0–6	0–60	Lower is better
Key secondary	Change from baseline in CGI-S at week 6	Clinical Global Impression – Severity	1–7	Lower is better
Key secondary	Response rate at week 6	≥50% reduction in MADRS from baseline	binary	Higher
Key secondary	Remission rate at week 6	MADRS total ≤10	binary	Higher
Secondary	Change in HAM-A, SDS, PHQ-9	anxiety, function, patient-reported	—	Lower
Secondary	CGI-I responder (score 1 or 2)	Global Impression – Improvement	1–7	Lower
Safety	TEAEs, C-SSRS, weight, sexual function, discontinuation-emergent symptoms	—	—	—

TRAP. "MADRS or HAM-D?" Know both and know why. HAM-D-17 is the historical standard and is still required by some regulators for comparability with legacy data; MADRS is more sensitive to change and does not carry the somatic-item baggage that inflates HAM-D scores for reasons unrelated to mood. Many programmes collect both, primary on one and supportive on the other. Saying "we

collected both and MADRS was primary because it is more responsive to change, with HAM-D-17 retained for historical comparability" is a complete, senior answer.

TRAP. CGI-S and CGI-I are different things and candidates conflate them. CGI-S is severity now, scored 1 to 7, and it has a baseline so change from baseline is meaningful. CGI-I is improvement relative to baseline, scored 1 to 7 with 4 as no change — it has **no baseline** and change from baseline is meaningless. CGI-I is analysed as a responder rate (score 1 or 2 = very much or much improved) or as an ordinal shift. Get this wrong and it is visible instantly.

Estimand

SAY THIS. "The primary estimand used a treatment-policy strategy for most intercurrent events but we had to be explicit about three: discontinuation of study drug, initiation of a prohibited rescue medication such as a benzodiazepine beyond protocol limits, and hospitalisation for worsening depression. The population was the modified intent-to-treat set — all randomised subjects who took at least one dose and had a baseline and at least one post-baseline MADRS. Summary measure: difference in LS mean change from baseline at week six."

SDTM — where psychiatric endpoints live

Domain	Content
QS	MADRS, HAM-D-17, HAM-A, CGI-S, CGI-I, PHQ-9, SDS — one record per item per visit
QS (`QSCAT='C-SSRS'`)	Columbia Suicide Severity Rating Scale — ideation and behaviour categories
AE	Adverse events including suicidal ideation or behaviour reported as AEs
CM	Concomitant psychotropics, rescue medication
EX, **DS**, **DM**, **VS**, **LB**	Standard; VS and LB matter for weight, lipids, prolactin in antipsychotic studies
SUPPQS / NSQS	Version of the instrument, rater identifier, and administration mode

SAY THIS. "One thing I always check in psychiatry is the rater. Under SDTM the rater goes to `QSEVAL`, and the rater identifier typically to a non-standard variable. In a MADRS study, rater consistency within subject across visits is a real quality signal — if a subject sees four different raters across six weeks, their score trajectory is partly measuring the raters. We produced a rater-consistency listing for the medical monitor, and it was one of the more valued outputs even though it was not in the SAP."

That is a strong answer because it is an output you added, not one you were told to build.

NOTE ON THE NEW NS– STRUCTURE. Under SDTMIG v4.0 the vertical `SUPPQS` is replaced by a horizontal `NSQS` dataset, one record per parent QS record with each non-standard variable as its own column carrying its own type, length and codelist. If you have implemented that migration, say so — it is current, it is not yet common, and it demonstrates you track the standard rather than inherit it.

ADaM

ADQS / ADEFF — the continuous parameters

Same BDS structure as CNS. The parameters:

PARAMCD	PARAM	Derived from
`MADRSTOT`	MADRS Total Score (0-60)	Sum of 10 item scores from QS
`MADRS01` … `MADRS10`	Individual items, retained for item-level analysis	QS direct
`CGISTOT`	CGI-Severity	QS direct
`CGIITOT`	CGI-Improvement (no baseline)	QS direct
`HAMA`	HAM-A Total	Sum
`SDSTOT`	Sheehan Disability Scale Total	Sum of 3 domains

```
/* MADRS total, with the scoring rule made explicit rather than assumed */
proc sql;
  create table madrs as
  select usubjid, visitnum, visit, qsdtc,
         sum(qsstresn)  as _sum,
         n(qsstresn)    as _nitems
  from  qs
  where qscat='MADRS' and qstestcd in ('MADRS01','MADRS02','MADRS03','MADRS04','MADRS05',
                                       'MADRS06','MADRS07','MADRS08','MADRS09','MADRS10')
  group by usubjid, visitnum, visit, qsdtc;
quit;

data adqs_madrs;
  set madrs;
  PARAMCD='MADRSTOT'; PARAM='MADRS Total Score (0-60)';
  /* SAP rule: total is missing if any item is missing. No prorating. */
  if _nitems = 10 then AVAL = _sum;
  else do; AVAL = .; end;
run;
```

TRAP. "Would you prorate?" Only if the scoring manual and the SAP say so, and MADRS conventionally does **not** — an incomplete MADRS is a missing MADRS. Prorating a clinician-rated scale invents a clinical judgement that was never made. If you prorate anywhere, `DTYPE` must be populated and the rule documented in the define. The wrong answer is "I would fill it in with the mean of the other items" offered as a default.

The responder parameters — where ADaM does real work

Response and remission are **DERIVED PARAMETERS IN THE SAME BDS DATASET**, not columns. This is a frequent ADaM misunderstanding and a good place to sound authoritative.

SAY THIS. "Response and remission are not flags hanging off the MADRS record. In ADaM they are their own parameters — one record per subject per visit, with `AVAL` of one or zero, and `CRIT1` carrying the human-readable criterion with `CRIT1FL` carrying its evaluation. That way the responder analysis is a straightforward subset by `PARAMCD`, the criterion is visible in the data rather than buried in a program, and define.xml can describe it."

```
data adqs_resp;
  set adqs(where=(PARAMCD='MADRSTOT' and ANL01FL='Y'));

  /* Response: >=50% reduction from baseline */
  PARAMCD='RESP50';
  PARAM  ='MADRS Response (>=50% reduction from baseline)';
  CRIT1  ='PCHG <= -50%';
  if n(PCHG) then do;
     if PCHG <= -50 then do; AVAL=1; AVALC='Responder';     CRIT1FL='Y'; end;
     else                do; AVAL=0; AVALC='Non-responder'; CRIT1FL='N'; end;
  end;
  output;

  /* Remission: MADRS total <=10 */
  PARAMCD='REMISS';
  PARAM  ='MADRS Remission (total <=10)';
  CRIT1  ='AVAL <= 10';
  if n(_aval_madrs) then do;
     if _aval_madrs <= 10 then do; AVAL=1; AVALC='Remitter';     CRIT1FL='Y'; end;
     else                      do; AVAL=0; AVALC='Non-remitter'; CRIT1FL='N'; end;
  end;
  output;
run;
```

TRAP. "What about a subject who dropped out before week six — are they a non-responder?" That depends entirely on the estimand, and the honest answer is that this is a statistical decision recorded in the SAP, not a programming default. Under a composite strategy, dropout is treated as non-response and the subject is a non-responder. Under a hypothetical or treatment-policy strategy with MAR, the record is missing and handled by the model or by imputation. Both are defensible; what is not defensible is doing one and documenting the other. Saying that sentence is the answer.

The analyses

Primary — MMRM, again

```
proc mixed data=adqs(where=(PARAMCD='MADRSTOT' and ANL01FL='Y'
                            and AVISITN>0 and MITTFL='Y')) method=reml;
  class USUBJID TRT01PN(ref='0') AVISITN POOLSITE RECURFL;
  model CHG = TRT01PN AVISITN TRT01PN*AVISITN
              BASE BASE*AVISITN POOLSITE RECURFL / ddfm=kr;
  repeated AVISITN / subject=USUBJID type=un;
  lsmeans TRT01PN*AVISITN / cl diff;
  ods output Diffs=dif LSMeans=lsm;
run;
```

TRAP — POOLED SITES. With four hundred and fifty subjects across seventy sites you cannot fit site as a class effect; many sites will have two or three subjects. The SAP defines a **pooling rule** — for example, sites with fewer than a threshold number of subjects per arm are pooled into a combined region — and that rule must be fixed **before unblinding**. If an interviewer asks how you handled site, the answer is "pooled by a pre-specified rule documented in the SAP and executed before unblinding, with the pooled assignment carried in ADSL so it is auditable." That is a much better answer than a statistical one, because it is the answer of someone who has been through a regulatory inspection.

Responder analysis

```
/* Logistic regression, stratified adjustment via covariates */
proc logistic data=adqs(where=(PARAMCD='RESP50' and AVISITN=6));
  class TRT01PN(ref='0') RECURFL POOLSITE / param=ref;
  model AVAL(event='1') = TRT01PN BASE RECURFL POOLSITE;
  oddsratio TRT01PN;
run;

/* Or CMH, if the SAP specifies a stratified rate difference */
proc freq data=adqs(where=(PARAMCD='RESP50' and AVISITN=6));
  tables RECURFL*TRT01PN*AVAL / cmh riskdiff(cl=newcombe);
run;
```

SAY THIS. "We reported the responder analysis both ways the SAP asked for — an odds ratio from logistic regression, and a stratified rate difference with a Newcombe interval — because clinicians read the rate difference and the number needed to treat, while the statistical test lives on the odds ratio. Producing both is not indecision; they answer different audiences."

Sensitivity and supportive analyses

Analysis	Purpose
ANCOVA on LOCF at week 6	Historical comparability; regulators still ask for it
ANCOVA on observed cases	Complete-case reference
Control-based multiple imputation	Tests the MAR assumption in the conservative direction
Tipping-point analysis	How bad would the unobserved data have to be to overturn the result
Analysis excluding a site with a data-integrity finding	Pre-specified only if the finding predates unblinding

TRAP. LOCF in depression behaves in the **opposite** direction from LOCF in dementia. In a six-week acute depression study, subjects generally improve over time, so carrying an early observation forward carries forward a worse score — which tends to be conservative for an effective drug. In progressive dementia, carrying forward freezes decline and is anti-conservative. Being able to say "LOCF is not inherently conservative; whether it is depends on the direction of the natural history" is a genuinely differentiating answer.

Safety — the psychiatry-specific outputs

C-SSRS IS THE ONE THEY WILL ASK ABOUT. The Columbia scale classifies suicidal ideation into five levels of increasing severity and suicidal behaviour into categories including preparatory acts, aborted attempt, interrupted attempt, actual attempt and completed suicide. The standard outputs are:

- Incidence of treatment-emergent suicidal ideation and treatment-emergent suicidal behaviour, by arm
- **Shift table** from worst baseline ideation category to worst post-baseline category
- Listing of every subject with any suicidal behaviour, with narrative cross-reference
- Time to first treatment-emergent suicidal ideation, where the programme requires it

SAY THIS. "Treatment-emergent here means a post-baseline category worse than the worst category at baseline — not simply the presence of any ideation post-baseline, because a substantial fraction of an

MDD population has ideation at baseline. Getting that definition right is the whole table. I have seen it programmed as 'any post-baseline ideation' and it materially overstates the signal."

Other psychiatry safety tables: weight and BMI change and the proportion with ≥7 percent weight gain; discontinuation-emergent signs and symptoms during taper; sexual dysfunction by ASEX or questionnaire; for antipsychotic programmes, extrapyramidal symptom scales — AIMS, Barnes Akathisia, Simpson-Angus — plus prolactin, fasting glucose, lipids and QTc.

The TLF inventory — psychiatry

Efficacy tables

#	Table
14.2.1	MMRM analysis of change from baseline in MADRS total by visit
14.2.2	MMRM analysis of change from baseline in CGI-S
14.2.3	Response rate (≥50% MADRS reduction) by visit, with odds ratio and rate difference
14.2.4	Remission rate (MADRS ≤10) by visit
14.2.5	CGI-I response rate and distribution of CGI-I scores at week 6
14.2.6	Change from baseline in HAM-A, SDS and PHQ-9
14.2.7	Sensitivity analyses of the primary — LOCF ANCOVA, observed case, control-based MI, tipping point
14.2.8	MADRS item-level change from baseline at week 6
14.2.9	Subgroup analyses — age, sex, baseline severity, recurrent versus single episode

Safety tables

TEAE overview; TEAE by SOC and PT; TEAE by severity and by relationship; SAEs; TEAEs leading to discontinuation; **C-SSRS TREATMENT-EMERGENT IDEATION AND BEHAVIOUR**; **C-SSRS shift table**; weight and BMI including the ≥7 percent category; vital signs and orthostatic changes; ECG including QTcF categorical outliers; clinical laboratory shifts; discontinuation-emergent symptoms.

Figures

#	Figure
14.2.1	LS mean change from baseline in MADRS total over time by arm, with SE bars — **the signature figure of a psychiatry programme**
14.2.2	Response and remission rates over time by arm
14.2.3	Cumulative distribution of percentage change in MADRS at week 6
14.2.4	Forest plot of subgroup LS mean differences
14.2.5	MADRS item-level effect sizes — radar or horizontal bar
14.2.6	Mean weight change over time

Listings

Individual MADRS scores by visit and item; subjects meeting response and remission with the criterion evaluated; all C-SSRS records for any subject with behaviour or worsened ideation; rescue medication use; rater assignment by subject and visit; protocol deviations affecting efficacy evaluability.

The read-out

SAY THIS. "The high dose separated from placebo — LS mean difference on MADRS of about minus 3.6 points at week six, p equal to 0.002 — and the low dose did not, at minus 1.8 and non-significant. Placebo change from baseline was about minus 11 points, which is a very typical placebo response and is exactly why the study needed the sample size it had. Response rate was 54 percent versus 39 percent, remission 32 percent versus 21 percent.

The programming point I would make is that the sensitivity analyses were what made the result believable rather than the primary. Dropout was 23 percent, and until the control-based multiple imputation and the tipping-point analysis showed the conclusion held under materially pessimistic assumptions about the missing data, we did not have a result we could defend. Delivering the primary was two days. Delivering the missing-data package that made the primary defensible was three weeks, and that is the part I would want to be judged on."

That last line is the closer. It reframes you from someone who runs procedures to someone who understands what makes a result survive review.

Rapid-fire drill — psychiatry

1. Why is CGI-I not analysed as change from baseline? — It has no baseline; it is already a comparison to baseline. Analyse it as a responder rate or ordinal shift.
2. What is a MADRS remitter? — Total score ≤10. Response is ≥50% reduction. They are different parameters, and a subject can be a responder without remitting.
3. How do you define treatment-emergent suicidal ideation? — Post-baseline category worse than the worst pre-existing baseline category, not merely any post-baseline ideation.
4. Why weekly visits? — Short study, high variance, and MMRM gains power from the intermediate visits; they also let you characterise onset of effect, which is a commercial as well as a clinical question.

5. Placebo response of 40 percent — is the study broken? — No, it is typical for MDD. It affects powering, not validity. The controls are enrichment strategies at design, centralised rater training, and duplicate independent rating at screening.
6. What is a pooled site and why does it matter? — Sites too small to fit as class levels are combined by a pre-specified rule fixed before unblinding; ADSL carries the pooled assignment.
7. Difference between the mITT and the safety set here? — Safety is anyone who took a dose; mITT adds the requirement of a baseline and at least one post-baseline efficacy measurement. Counts differ and the analysis-set table must reconcile them exactly.
8. A subject's week-4 MADRS is 62. — Impossible; the maximum is 60. That is a data-integrity finding raised as a query, not something you cap silently in ADaM. Never quietly repair an out-of-range value in a derived dataset.

The cross-cutting layer — what makes you sound principal-level

Everything in the three therapeutic-area modules is domain knowledge. This module is the layer that sits above all three, and it is where a panel decides whether you are a very good senior programmer or someone who can own a standard.

Estimands, said properly

ICH E9(R1) defines an estimand by five attributes. Learn them as a list you can recite, because "we used the ITT population" is not an estimand and a statistician will know that instantly.

Attribute	The question it answers	Oncology OS example
Treatment	Which regimens are being compared	Agent + doublet versus doublet
Population	Which subjects	All randomised (ITT)
Variable / endpoint	What is measured on each subject	Time from randomisation to death from any cause
Intercurrent events	What happens to events that occur after randomisation and affect interpretation, and the strategy for each	Discontinuation and subsequent therapy handled by treatment policy
Population-level summary	How the variable is summarised into a single comparison	Hazard ratio, and difference in medians

The five intercurrent-event strategies:

Strategy	What it does	Where you see it
Treatment policy	Ignore the event; use the outcome regardless	OS in oncology
Hypothetical	Estimate what would have happened had the event not occurred	"Had rescue medication not been available"
Composite	Fold the event into the endpoint itself	Dropout counted as non-response
While on treatment	Use only the outcome up to the event	Some safety and symptom endpoints
Principal stratum	Restrict to the stratum in whom the event would not occur	Rare; adherence-defined subpopulations

SAY THIS. "The practical consequence for me as a programmer is that the estimand determines the ADaM, not the other way round. A composite strategy means dropout becomes a value of `AVAL` — a non-responder record that exists. A hypothetical strategy means that same record is missing and handled by the model. Same collected data, different dataset, because the estimand differs. When a SAP describes an estimand loosely, that ambiguity lands on my desk as an unanswerable derivation question, so I read the estimand section first and I query it before I write anything."

That is one of the strongest paragraphs in this entire pack. It links a regulatory concept directly to a programming consequence, which almost no candidate does.

Analysis sets, and getting the counts to reconcile

Set	Definition	Where flagged
Screened	Signed consent	Usually outside ADSL
Randomised / ITT	All randomised, analysed as randomised	`ITTFL` / `RANDFL`
mITT	Randomised + dosed + baseline + ≥1 post-baseline	`MITTFL`
Safety	Received ≥1 dose, analysed as **treated**	`SAFFL`
Per-protocol	No major protocol deviations affecting the endpoint	`PPROTFL`
Efficacy-evaluable	TA-specific, e.g. ≥1 post-baseline tumour assessment	`EFFFL`
Responder subset	Achieved a response — for DOR	`RESPFL`

TRAP. "Analysed as randomised or as treated?" ITT is as randomised — `TRT01P`. Safety is as treated — `TRT01A`. When someone receives the wrong kit, they appear in the efficacy tables under their randomised arm and in the safety tables under what they actually got, and those two tables legitimately disagree on N. You must be able to explain that calmly, because a reviewer will ask why your safety table has 448 subjects and your efficacy table has 450.

SAY THIS. "The first output I build on any study is the disposition and analysis-set table, and the first QC I run is a reconciliation: randomised equals the sum of the arms, safety plus never- dosed equals randomised, and every exclusion from every set has a documented reason attached to the subject. If those do not reconcile at dry run, nothing downstream is trustworthy."

Multiplicity — the thing that decides whether a secondary endpoint counts

SAY THIS. "In all three studies the key secondary endpoints were tested in a pre-specified hierarchy — a fixed-sequence gatekeeping procedure. Testing stops at the first non-significant endpoint, and everything below it is reported descriptively without a claim. My job was to make the outputs honest about that: an endpoint below a broken gate does not get a p-value presented as if it were confirmatory. I have seen tables where it did, and it is the kind of thing that costs a label claim."

Know the vocabulary: family-wise error rate; fixed-sequence (hierarchical) testing; Hochberg and Holm procedures; graphical approaches with alpha propagation (Bretz–Maurer–Bornkamp); alpha-spending functions for group-sequential designs (O'Brien–Fleming and Pocock boundaries, and the Lan–DeMets spending approach that lets interim timing vary).

TRAP. "Your interim analysis crossed the boundary. What happens to the alpha for the final?" Alpha spent at the interim is not available at the final; the final boundary is adjusted accordingly, and that adjustment is pre-specified in the SAP and executed by the independent statistical group, not the study team. Knowing where the firewall sits — that the unblinded interim outputs are produced by an independent group and the study team stays blinded — is a governance point that plays very well at AD level.

The ADaM structure question

Panels routinely ask you to classify a dataset. Know the four:

Class	Structure	Examples
ADSL	One record per subject	Subject-level, always exactly one
BDS	One record per subject per parameter per analysis timepoint	ADQS, ADLB, ADVS, ADEFF, ADTR
BDS-TTE	BDS with time-to-event conventions — `CNSR`, `STARTDT`, `EVNTDESC`	ADTTE
OCCDS	One record per subject per occurrence	ADAE, ADCM, ADMH

TRAP. "Why is ADAE not BDS?" Because an adverse event is an occurrence, not a measurement of a parameter at a timepoint. It has no `PARAMCD` / `AVAL` pair in the BDS sense; it has occurrence flags (`AOCCFL`, `AOCCSFL`, `AOCCPFL`) that mark the first or worst occurrence for counting purposes. Explaining occurrence flags — that `AOCCPFL` marks one record per subject per preferred term so a subject is counted once in a PT row — shows you have built a safety table rather than just run one.

TRAP. "Can you have two records with the same `USUBJID`, `PARAMCD` and `AVISIT` in a BDS dataset?" Only if a further variable makes the key unique — typically `DTYPE`, so that an observed record and an LOCF-imputed record can coexist. Otherwise it is a define-level structure violation and Pinnacle 21 will flag it.

Traceability — the concept behind the whole standard

SAY THIS. "ADaM's core principle is that a reviewer can get from any analysis result back to the SDTM record it came from without asking me. That is what `SRCDOM`, `SRCVAR` and `SRCSEQ` are for at record level, and it is what intermediate parameters are for at derivation level. If a derivation has four steps, I would rather carry the intermediate as its own `PARAMCD` than bury it in a data step, because then the derivation is inspectable in the data instead of only in the code."

Two kinds of traceability, and knowing both terms is worth saying:

- **Metadata traceability** — the define.xml describes, for every variable, its origin, its derivation and its controlled terminology.
- **Data-point traceability** — the record itself points back to its source record.

Controlled terminology and define.xml

SAY THIS. "Every codelist in the study is versioned CDISC controlled terminology, and the version is stated in the define. When CT updates mid-study you do not silently adopt the new version — you assess the impact, decide with the standards group, and document. I built a CT compliance check that compares every coded variable in every domain against the declared CT package and reports extensible versus non-extensible violations separately, because an extension of an extensible codelist is allowed and an extension of a non-extensible one is a finding."

That is a genuine, specific, verifiable capability and it sets you apart from candidates who describe controlled terminology as a lookup table.

Define.xml deliverables you should be able to name: the define.xml itself, the stylesheet, the annotated CRF, the reviewer's guides (SDRG for SDTM and ADRG for ADaM), and — in the ADRG — the analysis-results metadata linking each key result to the dataset, the selection criteria and the programming statements that produced it.

SAY THIS ABOUT ARS. "The direction of travel is CDISC's Analysis Results Standard, which makes the analysis result itself a machine-readable object rather than a page in a reviewer's guide. I have been building toward that shape — outputs defined as metadata that a generator consumes — which is the same architectural idea as the metadata-driven specification work I did for the non-standard variable domains."

Validation and independent QC — the part that gets you the senior role

SAY THIS. "Every primary and key secondary output was double-programmed independently. Not reviewed — independently re-derived from SDTM by a second programmer working from the SAP and the shell, blind to the production code, and then compared. For the ADaM datasets that is a record-level compare on the analysis variables; for the TLFs it is a cell-by-cell comparison of the produced output, not an eyeball of the PDF.

The discipline point I would make is that a comparison that passes tells you nothing unless you can demonstrate that the two derivations were genuinely independent. If the QC programmer reads the production code first, the QC has been reduced to a proofread. That is a process control, not a technical one, and it is the one that actually fails in practice."

Then, if you have it, the concrete example:

SAY THIS. "I built an engine that does exactly that in production — the primary output is generated in one language and independently re-derived in a second, then diffed cell by cell with severity-ranked findings and human sign-off. On a live adverse-event table it surfaced four high-severity defects that had passed human review. I am careful about how I describe it: it is not an approved tool, and it does not make anything compliant by itself. What it produces is a sponsor-acceptable validation package. That distinction matters and I hold it firmly."

Twelve questions that cut across all three areas

1. **"Walk me from a protocol objective to a table number."** — Use the seven-link spine from the front matter. Do not skip links; the skipping is what they are testing.
2. **"What is the difference between ITT and mITT here, and who decided?"** — Definitions, plus: the statistician decided and it is in the SAP, and ADSL carries the flag so the choice is auditable.
3. **"A finding at dry run changes an ADaM derivation. What is your process?"** — Impact assessment across all downstream outputs, change to the spec first then the code, re-run, re-QC the affected outputs, document in the ADRG, and notify anyone holding an earlier version. Never patch the output.
4. **"How do you handle partial dates?"** — Imputation rules live in the SAP, are applied in ADaM not SDTM, and `*DTF` flags record what was imputed. SDTM keeps the character `--DTC` exactly as collected. Never impute into SDTM.
5. **"Treatment-emergent — define it."** — On or after first dose and on or before last dose plus the protocol-defined follow-up window. The window is protocol-specific, and events with fully missing start dates are conservatively treated as treatment-emergent unless the SAP says otherwise.
6. **"How would you QC a Kaplan–Meier figure?"** — Compare the underlying `ProductLimitEstimates` dataset, not the picture: number at risk at each tick, the median and its CI, the number of events and censors, and the step positions. Independently re-derive the ADTTE first, because a figure QC that starts at the figure is checking the plotting, not the analysis.

7. **"What goes in the SDRG versus the ADRG?"** — SDRG explains SDTM conformance issues, sponsor decisions and the mapping story; ADRG explains the analysis datasets, the derivations, the analysis-set definitions and the link from results to data.
8. **"Your Pinnacle 21 report has 300 findings."** — Triage: rejects and errors first, then warnings, then notices. Each one either gets fixed or gets an explanation in the reviewer's guide. A finding with a documented sponsor rationale is acceptable; an undocumented one is not. The count itself is not the metric.
9. **"How do you version and control your programs?"** — Version control with a controlled release branch, no direct edits to production, a documented promotion path from development to validated, and a run log that ties every output to the exact program version and dataset version that made it. Under 21 CFR Part 11, who ran what and when has to be reconstructible.
10. **"What do you do when the statistician's SAP text is ambiguous?"** — Raise it in writing before programming, get the clarification recorded as a SAP amendment or a documented decision, and only then implement. Do not resolve ambiguity by choosing; that puts a clinical decision in a program.
11. **"How do you use AI tooling in a regulated deliverable?"** — Generation is assisted; acceptance is not. Anything generated is treated as a draft artifact that must pass independent re-derivation and human sign-off before it is used, and the provenance of what was generated is recorded. Nothing produced this way is an approved output; the deliverable is a validation package a sponsor can accept.
12. **"What would you change about how the industry does this?"** — Have a real answer. Ours: the specification and the program should be generated from one metadata source so they cannot drift, and the QC should be an independent re-derivation by construction rather than a review step that can be skipped under timeline pressure.

The sentence to keep in reserve

If a panel presses on something you have not done, the answer is not to bluff. It is:

> "I have not run that specific analysis. Here is the closest thing I have built, here is the part that transfers directly, and here is the part I would need to learn — which is the model specification, not the data flow."

Panels at this level hire on judgement, and the willingness to draw that line accurately is the single clearest signal of it.

The performance — scripts, drills and recovery

The previous modules are what you know. This one is what you say. Read these aloud. Material you have only read silently collapses under interview pressure; material you have spoken three times does not.

The ninety-second opener — use this whichever area they name

> "Let me take one study and go end to end, because I think the useful thing is the chain rather than any single piece of it.
>
> [One sentence: phase, indication, design, N, duration.]
>
> The primary objective was [X]. That translated into a primary endpoint of [Y], and the estimand treated [the main intercurrent event] with a [strategy] strategy — which matters, because it determined how I built the analysis dataset.
>
> The endpoint was collected as [CRF description], which lands in SDTM domains [A, B and C]. I derived it into [ADaM dataset], where the parameter is [PARAMCD], the analysis value is [AVAL definition], and the key derived variables are [two or three].
>
> The SAP called for [analysis], which is [PROC] with [the one option worth naming]. That produced table [number] and figure [number].
>
> The study [read out]. Happy to go deeper at any point in that chain — the interesting programming was in [the one genuinely hard part]."

That last line is the control lever. It invites the panel into the part of the chain where you are strongest, and most panels take the invitation.

Full talk track — oncology, forty minutes

0–3 MIN. SET THE STUDY. Phase 3, first-line advanced NSCLC, ~600 subjects, 1:1, agent plus platinum doublet versus doublet, stratified by PD-L1, histology, region. Treat to progression, then survival follow-up every twelve weeks. Two interims with an O'Brien–Fleming spending function.

3–8 MIN. OBJECTIVES, ENDPOINTS, ESTIMAND. Primary OS. Key secondaries PFS by BICR, ORR. Then the estimand paragraph from Module 1 — treatment policy for OS, and why PFS needs the censoring sensitivity set instead.

8–15 MIN. DATA FLOW. The tumour trio: TU identifies the lesions, TR carries the measurements, RS carries the response. Deaths from DM and DS. Subsequent therapy from CM and PR. Then the last-known-alive derivation — this is your first depth moment, so slow down here.

15–25 MIN. ADAM. ADSL first: randomisation date, death date, last known alive, both stratum versions, population flags. Then ADTTE: the parameter table, the OS derivation, the PFS derivation with its censoring rules, `CNSR` orientation. Then ADRS: OVRLRESP, BOR, CBOR and the confirmation look-ahead. Then ADTR and the nadir-based progression rule with its 20 percent and 5 mm criteria.

25–33 MIN. ANALYSIS. `PROC LIFETEST` with the stratified log-rank via `strata / group=`; the log–log confidence type; the number-at-risk row. `PROC PHREG` with `ties=efron` and stratification. `PROC FREQ` with exact binomial within arm and CMH between arms. Then why DOR carries no p-value.

33–40 MIN. OUTPUTS AND READ-OUT. The seven efficacy tables by number, the safety block, the seven figures. Then the read-out paragraph, then the honest caveat about subsequent therapy diluting OS, then the process improvement — parameterising the censoring rules.

Full talk track — CNS, forty minutes

0–3. Phase 3, early symptomatic Alzheimer's, amyloid-confirmed, ~1700 subjects, 1:1, IV every four weeks, 78 weeks, stratified by APOE ε4 and symptomatic AD medication use.

3–8. Primary CDR-SB change at week 78. Key secondaries ADAS-Cog13, ADCS-ADL, amyloid PET. Then the direction-of-scale point — some go up with worsening, some down — and how you protected against it in metadata rather than in a programmer's memory.

8–15. QS structure: `QSCAT`, `QSTESTCD`, `QSSTRESN`, `QSEVAL`. The informant-based scales and study-partner continuity. Where biomarkers and imaging live, and the honest note that imaging quantification mapping is not uniform across sponsors.

15–25. ADQS as BDS. `PARAMCD` list. Baseline definition. Then visit windowing at length — this is your depth moment in CNS, and the tie-break rule and `ANL01FL` are the substance. Then total-score scoring rules and missing items.

25–33. MMRM. The model statement line by line, `type=un`, the convergence step-down sequence, Kenward-Roger, `repeated` not `random`, and why not LOCF in a progressive disease. Then missing data: control-based multiple imputation, tipping point, `PROC MI` / `PROC MIANALYZE`, the fixed seed.

33–40. Tables, the ARIA safety block by APOE genotype, the figures — and spend your time on the CDF plot and why you would rather present the distribution than the mean alone. Read-out with the honest three-part framing: biomarker decisive, clinical effect real but small and debated, safety manageable but genotype-dependent.

Full talk track — psychiatry, forty minutes

0–3. Phase 3 MDD, ~450 subjects, two doses versus placebo, six weeks double-blind, weekly visits, MADRS ≥26 and CGI-S ≥4 at entry.

3–8. Primary MADRS change at week 6. Key secondaries CGI-S, response, remission. Then the noise paragraph — placebo response, dropout, rater variance — and how every design and analysis choice traces back to it.

8–15. QS structure, MADRS items, the CGI-S versus CGI-I distinction, C-SSRS, rater identity and the rater-consistency listing you added on your own initiative. Mention NS– if you have implemented it.

15–25. ADQS: MADRSTOT and the no-prorating scoring rule. Then responder parameters as their own `PARAMCD` with `CRIT1` and `CRIT1FL` — this is your depth moment in psychiatry, because most candidates describe them as flags. Then the estimand-dependent handling of dropouts as non-responders.

25–33. MMRM again, plus the pooled-site rule fixed before unblinding. Logistic regression and CMH for responders, and why you present both an odds ratio and a rate difference. Then the sensitivity suite and the LOCF direction argument — conservative here, anti-conservative in dementia.

33–40. Tables, the C-SSRS block including the treatment-emergent definition, the figures. Read-out with the closer: "delivering the primary was two days; delivering the missing-data package that made it defensible was three weeks."

Panel drill — thirty questions, answer out loud

DESIGN AND ENDPOINTS

1. Why is OS the gold-standard oncology endpoint despite needing the longest follow-up?
2. When is PFS acceptable as a primary endpoint for approval?
3. What is the difference between ORR and DCR, and why is DCR weaker evidence?
4. Your primary is continuous and measured at six visits. Why analyse all six rather than only the last?
5. What is a landmark analysis and what is its main weakness?
6. Define the estimand for your primary endpoint using all five attributes.

SDTM

1. Which domains would you need to derive PFS, and what does each contribute?
2. How do TU, TR and RS relate to one another?
3. Where does a total questionnaire score live if the EDC calculated it?
4. What is the difference between `--ORRES`, `--STRESC` and `--STRESN`?
5. Why can you never impute a date in SDTM?
6. Under SDTMIG v4.0, what replaced SUPPQUAL and what does that change for you?

ADAM

1. Classify ADAE, ADTTE, ADQS and ADSL by structure and justify each.
2. What does `CNSR = 0` mean?
3. When must `DTYPE` be populated?
4. What is `ANL01FL` for, and why not simply subset the dataset?
5. How do you make a responder definition traceable?
6. Two records share `USUBJID`, `PARAMCD` and `AVISIT` — is that legal?
7. What is `AOCCPFL` and what would go wrong without it?
8. How do you record traceability from an analysis value back to SDTM?

ANALYSIS

1. Stratified versus unstratified log-rank — how do you get each in SAS?
2. Why Efron rather than Breslow for ties?
3. Why Kenward-Roger?
4. Your unstructured MMRM does not converge. What now?
5. Why negative binomial rather than Poisson for relapse counts?
6. How many imputations, and why does the seed matter?
7. Log-rank and Cox disagree slightly — which do you report?

GOVERNANCE

1. How do you keep the study team blinded through an interim analysis?
2. Your key secondary is significant but the gate above it failed. What appears in the table?
3. How do you validate an output produced with AI assistance?

Recovery lines — for when you are caught

Situation	What to say
You do not know	"I have not built that. The nearest thing I have done is [X], and the part that transfers is the data flow; what I would need to learn is the model specification."
You gave a wrong number	"Let me correct that — I said [X], it is [Y]. I would rather fix it now than have you find it later."
They challenge a choice	"That is a fair challenge. The SAP specified [A]; the argument for [B] is [reason]. If I were writing it now I would want [the deciding consideration] settled first."
You lose your thread	"Let me come back up a level — where I was in the chain is [link]. The next step is [link]."
They ask something outside programming	"Outside my lane, but the way it reached me was [the programming consequence], so here is how I saw it from where I sat."

The forty-eight hours before

- Read the three talk tracks aloud, timed. If any runs long, cut the middle, never the read-out.
- Rehearse the seven-link spine until you can recite it as a list.
- Pick **one** number per study you are certain of, and let the rest be approximate. A confidently wrong hazard ratio is worse than "roughly 0.7".
- Prepare two questions for them that only someone who has done the work would ask. For example: "How do you handle the firewall between the independent statistical group and the study team at interim?" and "Where does the analysis-results metadata live in your process today — reviewer's guide, or something machine-readable?"
- Have one failure story ready with a real process change attached to it. Every panel asks, and the candidates who have not prepared one give an answer that sounds invented, because it is.

The one thing to remember walking in

The panel is not checking whether you know what `PROC LIFETEST` does. They are checking whether, when a protocol objective lands on your desk, you know what has to be true of the data, the dataset, the model and the output for that objective to be answerable — and whether you would say so when it is not.

Talk about the chain. Stay honest about the parts you did not own. Close on what you changed.